

Lecture Notes on Data Engineering
and Communications Technologies 123

Zheng Xu · Saed Alrabaee ·
Octavio Loyola-González · Xiaolu Zhang ·
Niken Dwi Wahyu Cahyani ·
Nurul Hidayah Ab Rahman *Editors*

Cyber Security Intelligence and Analytics

The 4th International Conference
on Cyber Security Intelligence and
Analytics (CSIA 2022), Volume 1

Lecture Notes on Data Engineering and Communications Technologies

Volume 123

Series Editor

Fatos Xhafa, Technical University of Catalonia, Barcelona, Spain

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at <https://link.springer.com/bookseries/15362>

Zheng Xu · Saed Alrabaee ·
Octavio Loyola-González ·
Xiaolu Zhang · Niken Dwi Wahyu Cahyani ·
Nurul Hidayah Ab Rahman
Editors

Cyber Security Intelligence and Analytics

The 4th International Conference on Cyber
Security Intelligence and Analytics
(CSIA 2022), Volume 1

 Springer

Editors

Zheng Xu
School of Computer
and Information Engineering
Shanghai Polytechnic University
Shanghai, China

Octavio Loyola-González
Altair Management Consultants
Madrid, Spain

Niken Dwi Wahyu Cahyani
Informatics Faculty
Telkom University
Bandung, Indonesia

Saed Alrabaei
Department of Information Systems
and Security
College of IT, United Arab Emirates
University (UAEU)
Abu Dhabi, United Arab Emirates

Xiaolu Zhang
Department of Info Systems &
Cyber Security
University of Texas at San Antonio
San Antonio, TX, USA

Nurul Hidayah Ab Rahman
Faculty of Computer Science and
Information Technology
Universiti Tun Hussein Onn Malaysia
Parit Raja, Johor, Malaysia

ISSN 2367-4512

ISSN 2367-4520 (electronic)

Lecture Notes on Data Engineering and Communications Technologies

ISBN 978-3-030-96907-3

ISBN 978-3-030-96908-0 (eBook)

<https://doi.org/10.1007/978-3-030-96908-0>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

The 4th International Conference on Cyber Security Intelligence and Analytics (CSIA 2022) is an international conference dedicated to promoting novel theoretical and applied research advances in the interdisciplinary agenda of cyber security, particularly focusing on threat intelligence and analytics and countering cybercrime. Cyber security experts, including those in data analytics, incident response, and digital forensics, need to be able to rapidly detect, analyze, and defend against a diverse range of cyber threats in near real-time conditions. For example, when a significant amount of data is collected from or generated by different security monitoring solutions, intelligent and next generation big data analytical techniques are necessary to mine, interpret, and extract knowledge of these (big) data. Cyber threat intelligence and analytics are among the fastest-growing interdisciplinary fields of research bringing together researchers from different fields such as digital forensics, political and security studies, criminology, cyber security, big data analytics, and machine learning to detect, contain, and mitigate advanced persistent threats and fight against organized cybercrimes.

CSIA 2022, building on the previous successful online meeting (2021 and 2020 due to COVID-19), in Wuhu, China, (2019) is proud to be in the fourth consecutive conference year. Each paper was reviewed by at least two independent experts. The conference would not have been a reality without the contributions of the authors. We sincerely thank all the authors for their valuable contributions. We would like to express our appreciation to all members of the program committee for their valuable efforts in the review process that helped us to guarantee the highest quality of the selected papers for the conference.

Especially, CSIA 2022 has a significant amount of international contributions, and authors are from China, India, USA, Philippines, UAE, Egypt, Turkey, Australia, Korea, Canada, Brazil, UK, etc.

Our special thanks are due also to the editors of Springer book series “Lecture Notes on Data Engineering and Communications Technologies,” Thomas Ditzinger, Suresh Dharmalingam for their assistance throughout the publication process.

Organization

Steering Committee Chair

Kim-Kwang Raymond Choo University of Texas at San Antonio, USA

General Chair

Ishfaq Ahmad University of Texas at Arlington, Arlington,
Texas, USA

Program Committee Chairs

Zheng Xu Shanghai Polytechnic University, China
Saed Alrabaee United Arab Emirate University, UAE
Octavio Loyola-González Altair Management Consultants, Spain
Xiaolu Zhang University of Texas at San Antonio, USA

Program Committee Vice-chairs

Niken Dwi Wahyu Cahyani Telkom University, Indonesia
Nurul Hidayah Ab Rahman Universiti Tun Hussein Onn, Malaysia

Publicity Chair

Neil. Y. Yen University of Aizu, Japan
Vijayan Sugumaran Oakland University, USA
Junchi Yan Shanghai Jiaotong University, China

Local Organizing Chairs

Chuang Ma	Shanghai Polytechnic University, China
Chun Pan	Shanghai Polytechnic University, China

Program Committee Members

Neil Yen	University of Aizu, Japan
Shunxiang Zhang	Anhui University of Science and Technology, China
Guangli Zhu	Anhui University of Science and Technology, China
Tao Liao	Anhui University of Science and Technology, China
Xiaobo Yin	Anhui University of Science and Technology, China
Xiangfeng Luo	Shanghai University, China
Xiao Wei	Shanghai University, China
Huan Du	Shanghai University, China
Zhiguo Yan	Fudan University, China
Zhiming Ding	Beijing University of Technology, China
Jianhui Li	Computer Network Information Center, Chinese Academy of Sciences, China
Yi Liu	Tsinghua University, China
Kuien Liu	Pivotal Inc, USA
Feng Lu	Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, China
Wei Xu	Renmin University of China, China
Ming Hu	Shanghai University, China
Abdelrahman Abouarqoub	Middle East University, Jordan
Sana Belguith	University of Auckland, New Zealand
Rozita Dara	University of Guelph, Canada
Reza Esmaili	Amsterdam University of Applied Science. Netherlands
Ibrahim Ghafir	Loughborough University, UK
Fadi Hamad	Israa University, Jordan
Sajjad Homayoun	Shiraz University of Technology, Iran
Nesrine Kaaniche	Telecom SudParis, France
Steven Moffat	Manchester Metropolitan University, UK
Charlie Obimbo	University of Guelph, Canada
Jibrán Saleem	Manchester Metropolitan University, UK
Maryam Shahpasand	Asia Pacific University Malaysia, Malaysia
Steven Walker-Roberts	Manchester Metropolitan University, UK

The 4th International Conference on Cyber Security Intelligence and Analytics

(CSIA 2022)

Program Book

Online Conference, March 19, 2022

Due to COVID-19, authors, keynote speakers, and PC committees will attend the conference online by Tencent Meeting.

<http://www.csia.fit/csia2022>

Conference Program at a Glance

Saturday, March 19, 2022, Online Tencent Meeting		
9:50–10:00	Opening Ceremony	Kim-Kwang Raymond Choo
10:00–10:40	Keynote 1	Octavio Loyola-González
10:40–11:20	Keynote 2	Xiaolu Zhang
11:30	Best Paper Awards	Zheng Xu
14:00–18:00	Technical tracks 1: Cyber intelligence and analytics from different layers of network space solutions	Xiaolu Zhang
	Technical tracks 2: Control systems, system integration, and industrial artificial intelligence	Octavio Loyola-González
	Technical tracks 3: Cloud computing, big data, the Internet of Things, and their system Intelligence	Saed Saleh Al Rabae
	Technical tracks 4: Machine learning, computer vision, image processing, and their intelligence applications	Niken Dwi Wahyu Cahyani
	Technical tracks 5: Computational intelligence, blockchain, VR, and their engineering intelligence applications	Nurul Hidayah Ab Rahman
	Technical tracks 6: Marketing decision making, E-commerce, and their sustainable intelligence computing	Reza Meimandi Parizi
	Technical tracks 7: Cyber intelligence for education course and applications	Zheng Xu

CSIA 2022 Keynotes

A Review of Fuzzy and Pattern-Based Approaches for Class Imbalance Problems

Dr. Octavio Loyola-González

Altair Management Consultants, Spain



Octavio Loyola-González received the Ph.D. degree in computer science from the National Institute for Astrophysics, Optics and Electronics, Mexico, in 2017. He worked as a distinguished professor and a researcher with the Tecnológico de Monterrey at Puebla, for undergraduate and graduate programs of computer sciences. He is currently responsible for running machine learning and artificial intelligence practice inside Altair Management Consultants Corporation, where he is involved in the development and implementation using analytics and data mining with the Altair Compass Department. He has outstanding experience in the fields of big data and pattern recognition, cloud computing, the IoT, and analytical tools to apply them in sectors where he has worked for as banking and insurance, retail, oil and gas, agriculture, cybersecurity, biotechnology, and dactyloscopy. From these applied projects, he has published several books and articles in well-known journals, and he has several ongoing patents as the manager and a researcher in Altair Compass. He is a member of the National System of Researchers in Mexico (ranked one). He received several awards from different institutions due to his research work on applied projects.

Non-negative Matrix Factorization-Based Heterogeneous Graph Embedding Method for Trigger-Action Programming in IoT

Dr. Xiaolu Zhang

University of Texas at San Antonio, USA



Xiaolu Zhang received the Ph.D. degree in computer science from Jilin University, Changchun, China, in 2016. He is currently an associate professor in the Department of Information Systems and Cyber Security, University of Texas at San Antonio (UTSA), San Antonio, TX, USA. Prior to his current position, he was a visiting Ph.D. student in the University of New Haven, West Haven, CT, USA, and a postdoc researcher with the University of Texas at San Antonio. His research interests include wide field of mobile, IoT, and computer security and forensics. Dr. Zhang was the recipient of a China Scholarship Council Scholarship for his doctoral work and UTSA's Endowed 1969 Commemorative Award for Teaching Excellence in 2021.

Oral Presentation Instructions

1. Timing: a maximum of 10 minutes total, including speaking time and discussion. Please make sure your presentation is well timed. Please keep in mind that the program is full and that the speaker after you would like their allocated time available to them.
2. You can use CD or USB flash drive (memory stick), make sure you scanned viruses in your own computer. Each speaker is required to meet her/his session chair in the corresponding session rooms 10 minutes before the session starts and copy the slide file (PPT or PDF) to the computer.
3. It is suggested that you email a copy of your presentation to your personal inbox as a backup. If for some reason the files cannot be accessed from your flash drive, you will be able to download them to the computer from your email.
4. Please note that each session room will be equipped with a LCD projector, screen, point device, microphone, and a laptop with general presentation software such as Microsoft PowerPoint and Adobe Reader. Please make sure that your files are compatible and readable with our operation system by using commonly used fonts and symbols. If you plan to use your own computer, please try the connection and make sure it works before your presentation.
5. Movies: If your PowerPoint files contain movies, please make sure that they are well formatted and connected to the main files.

Short Paper Presentation Instructions

1. Maximum poster size is 0.8 meter wide by 1 meter high.
2. Posters are required to be condensed and attractive. The characters should be large enough so that they are visible from 1 meter apart.
3. Please note that during your short paper session, the author should stay by your short paper to explain and discuss your paper with visiting delegates.

Registration

Since CSIA 2022, use online meeting, no registration fee.

Contents

Cyber Intelligence and Analytics from Different Layers of Data Warehouse Solutions	
Intelligent Archives Warehouse Management Based on New Information Technology	3
Mengxue Zhang and Rasha Almajed	
Design and Algorithm of File Management System Based on RFID . . .	11
Weimeng Liu	
Design of Information System for Accounting File Management Based on Computer Technology	19
Yan Sun	
Substation Cable Laying Schedule Based on GIM 3D Model	28
Hailiang Wu, Lin Xing, Nan Cheng, Mingfu Li, and Jizhi Su	
Social Service Management System Based on Logistic Regression Algorithm	37
Huajun Li	
Timing System of Track and Field Competition Based on Data Analysis Algorithm	45
Yingyue Hu and Sinem Alturjman	
Financial Software Perception Function Based on Big Data Intelligent Technology	52
Hongyan Li	
Effective Combination of Intelligent Information Technology and Labor Education Data Analysis Algorithms	61
Shasha Lv	

Chinese Language and Literature Relational Database Mining Based on Association Rules Algorithm 70
 Yingxia Tang

Internet Financial News Text Classification Algorithm Based on Blockchain Technology 78
 Long Chen

Measurement and Change Analysis of Trade Implied Carbon Based on Input-Output Method and SDA Method Under Modern Information Technology 86
 Jing Ye and Shalini Kumar Shukla

The Application of Big Data in Network Security and Intelligence Analysis 96
 Wang Zheng and Juan Liu

Macroeconomic Forecast Based on Neural Network Algorithm and Genetic Algorithm 103
 Yuanjie Wang

Intelligent Optical Distribution Technology in Power Communication Transmission 112
 Xu Zhang, Xin Xiao, Le Ma, Chao Liu, Xiangju Sun, and Xiaowei Cheng

Intelligent Puncture Equipment in Sampling of Gynecological Tumors 120
 Hui Zhu

Design of Environmental Art Optimization System Based on Improved Particle Swarm Optimization Algorithm 128
 Yi Fu, Chunjiang Yu, and Saad Metawa

Software-Assisted Japanese Phonetic Teaching Based on Phonetic Visualization 135
 Xinyi Zhang and Lin Liu

Internet Financial Risk Forecast System Based on Artificial Intelligence Algorithm 143
 Lianying Ge, Yachen Shen, and K. Vijayakumar

Structure Analysis of Digital Economy Association Network Based on Intelligent Algorithm 151
 Jing Tan and Yao Gu

The Digital New Infrastructure Construction System of Provincial Power Grid 158
 Xiao Lu, Lei Wei, and Zhe Wang

Application of Digital New Infrastructure Business Scenarios in Power Grid Enterprises 166
 Qinwei Dong, Chengling Jiang, and Zhoujun Ma

Risks and Countermeasures of Cyber Security Enterprises 175
 Jianwei Ma, Zhuojun Yao, and Sinem Alturjman

Environmental Economic Decision Support Information System Based on GIS 184
 Yuqi Wang

Network Security Management and Protection in the Context of Emerging Technologies 192
 Qian Li

Network Security Design of E-commerce Sales Management System Based on Neural Network Algorithm 199
 Shaocheng Zhang, Jiali Yang, Tingwei Wang, and Cong Wang

Data Mining and Computer Software Database Programming Technology 208
 Xiaoqing Wang

Design of Visualization System Under Centralized Monitoring of Regional Power Grid Based on Data Mining Algorithm 216
 Piao Liu, Jinman Lou, Kai Yu, Xiaoxia Li, and Youjun Feng

Application of Data Mining Algorithm in Realizing Monitoring and Visualization of Transmission Network Government Affairs 225
 Fangtao Wan, Shengxiang Xie, Jianshu Yuan, and Lixia Zheng

Cyber Security Practice in Organizations Based on Internet Environment 233
 Jing Lin

Analysis of the Impact of New Energy Access on Smart Distribution Network Operation Based on Deep Learning Technology 240
 Guihua Qiu, Zhirui Tang, and Zijia Kuang

Application of “Blockchain + 5G” Technology in the Transformation of Financial Digital Intelligence 248
 Yanqing Chen

Comprehensive Evaluation Model of TOPSIS Method Based on Data Processing Technology 256
 Yi Huang and Wenbin Liu

Data Mining Algorithm of Hotel Customer Data Based on Mobile Computing 263
 Jun Wang

Design and Development of Interior Design CAD System Based on Data Mining Algorithm 270
 Yun Peng and Feilu Hang

Design and Research of Public Rental Housing Rent Pricing System Based on Mobile Internet 278
 Qiong Zhang and Hui-yong Guo

Design of Interior Design Database Based on Ant Colony Algorithm 285
 Yuhan Zhang

CPS Systems, System Integration, and Real-World Applications

Application of Computer BIM Software Technology in Building Information Model 295
 Xiu Luo and Amar Jain

The Design of Mental Health Information Data Management System Under the Background of Informationization 303
 Nan Jiang and Rajan Singh Tiwari

Design of Rock Climbing Data Acquisition System Based on LoRa 311
 Dan Zhang, Jing Huang, Luyuan Cai, and Shuai Cao

Research on Realization of Yi Speech Synthesis in Robot 318
 Yihong Jiejue and Saad Metawa

Application of Deep Learning Algorithm in National Fitness Monitoring 327
 Xiaoliang Li and Liping Zhang

The Optimization Algorithm of Tourism Route Based on Deep Learning Theory 335
 Jia Xue

Construction of an Intelligent Student Management and Evaluation Information System Based on Fuzzy Neural Network 343
 Cheng Qian

Forecasting Model of Number of Entrepreneurs in Colleges and Universities Based on PSO Algorithm 351
 Pingping Yang

Design of Automatic Evaluation of Machine English Translation Based on BP Neural Network 359
 Yan Zhang

Design and Implementation of Multimedia Vocal Music Learning System Based on Network Information Technology 368
 Fang Hui

Sports Quality Test Recognition System Based on Fuzzy Clustering Algorithm 376
Dayang Ming

Conceptual Model of Effectiveness Evaluation of Foreign Language Courses Based on K-means Mean Algorithm 384
Guilan Chen

The Acquisition System of Periodontitis Alveolar Bone Structure Based on Digital X-ray Images 392
Weicheng Yan and Liyun Dai

Substation Construction Management and Control System Based on 3D Modeling Technology 400
Hongwei Yang, Wuchen Zhang, Yamin Wang, Jizhi Su, and Nan Cheng

Design and Implementation of Oral English Test System Based on Data Mining Algorithm 408
Xiaolan Guo and Neha Jain

Design of Intelligent Recognition System of Employees' Behavior in Coal Mine Based on VR Technology 416
Zenghui Liu

Design of Chinese Linguistics Teaching System Based on K-means Clustering Algorithm 424
Hongling Ma

Construction of Maker Educational Resource Sharing Platform Based on Web Technology 432
Xia Li

Travel Path Planning Algorithm Based on Deep Learning Technology 440
Chenglong Yu

Risk Forecast of International Engineering Project Based on Data Mining Technology 448
Depeng Li, Wei Liu, and Rasha Almajed

Speech Recognition Technology of English Translation System Based on DTW Algorithm 455
Peilan Peng

Design and Implementation of the Life Cycle Management System of Power Tools Based on the Internet of Things 462
Zhengfeng Xie, Hongyun Shi, Lianggang Xu, Shichun Wang, Shi Qiu, Long Yang, and Rajan Singh Tiwari

Building Information Model Monitoring System Based on BIM and Optimized Ant Colony Algorithm 471
Xiao Cai, Lei Pan, Jiawen Gao, Jinbang Lin, Shilong Zang, and Tianhao Zhang

Development Status of High-End Equipment Manufacturing Industry Based on Neural Network and Tobit Panel Regression Model 480
Zilong Di

Construction of Financial Credit Risk Evaluation System Model Based on Analytic Hierarchy Process 488
Shenghua Qian

Product Intelligent Design Model Based on Genetic Optimization Algorithm 497
Tengjiao Liu and Shalini Kumar Shukla

Advanced Consumer Psychology Based on Decision Tree Data Mining Algorithm 505
Yue Lu and Jie Zhou

Design and Implementation of Hotel Management Information System Based on Intelligent Terminal 513
Yanjie Yu

Fuzzy Sets Qualitative Comparative Analysis (fsQCA) of Helicopter Accident Casualties Based on HFACS Model 521
Bei Zhou

Animation Game Database Model Based on Virtual Reality Technology 530
Huihuang Wu

The Application of Drones in Urban Fire Monitoring 537
Gang Wang, Yang Cao, and Jun Zhou

Auto Tailgate Automatic Control System Based on Single Chip Microcomputer 545
Ying Zhao

Corpus-Based Error Analysis of Students’ Online Writings in Automated Essay Scoring System 554
Ronggen Zhang

Design and Development of Public Art Design System Based on VR Intelligence 562
Zhiyuan Niu and Shishi Xiang

Research on Logistics Management Information System of Electronic Commerce Based on Computer Information Technology 569
Yang Ding and Neha Jain

AHP-Based Small Cross-Border e-Commerce Export Trade Logistics Model Selection and Application Analysis 577
 Jiang Yu

Design and Implementation of RFID-Based Logistics Clearance Information Platform 585
 Qiong Yang

Machine English Translation Scoring System Based on Deep Neural Network in the Internet of Things Environment 593
 Longyuan Xiao

System Design of Single On-line Uninterruptible Power Supply 601
 Xifeng Shen, Ming Liu, Hua Liu, Jinfeng Dong, Guoqing Jiang, and Xin Zhao

Fault and Detection of Electrical Control System Based on Intelligent Algorithm 609
 Ling Liu

Design and Implementation of Graduate Student Enrollment Management Information System Based on Rsync Algorithm 617
 Cui Liu

Information System of Constructive English Teaching Platform Based on RBF Algorithm 626
 Jing Gao

Dynamic Grouping Design System of Leisure Sports Based on Computer Intelligent Algorithm 634
 Lipeng Zhang

The Frame Display of the Digital Clothing Display Design System Based on Intelligent Man-Machine Cooperation 641
 Le Liu

Cloud Computing, Big Data, the Internet of Things, and Their System Intelligence

Classification and Recognition of Internet Literature Text Based on Data Mining Technology 651
 Rong Xiong, Zichen Li, Yongjun Qi, and Thomas Lavoie

Design and Realization of Applied Linguistics MOOC System Based on Feature Extraction Algorithm 660
 Hongyan Sang

Characteristics of Piano MOOC Learning Needs Based on Data Mining Technology 669
 Yaokun Yang and Allan Comstock

Application of Grey Model Algorithm in Urban GDP Growth Calculation Model System 676
Chenhe Wang and Arnold Salave

Design and Implementation of Nursing Big Data Platform Based on Data Mining Technology 684
Dapeng Liu, Hui Zheng, Qionqiong Yan, Ying Li, and Cuicui Huang

Construction of the Intelligent Online Tourism English Model Corpus Based on ESP 693
Jianzhou Cui

Data Mining of E-Commerce Enterprise User Preferences in the Context of Big Data 700
Lielong Yang and Xingtian Yin

Analysis of English Teaching Data on Cloud Platform Based on Data Mining Technology 708
Zhuolin Zhou

Realization of the Character Analysis Algorithm of Science Fiction Characters Based on Text Information 715
Ying He

Application of Mass Data Retrieval Technology Based on Cloud Computing 723
Ruijuan Li and Shunji Zhang

Data Analysis of Pharmacy Training Management System Based on Apriori Algorithm 731
Hui Zheng, Dapeng Liu, Jing Zhou, Teng Ma, and Yan Wang

Health Qigong Sports Network Learning Platform in Cloud Computing 739
Zhiguo Tian and Wei Diao

Construction of Stadium Information Management Platform Based on Cluster Intelligent Algorithm 746
Jing Yang

Disclosure and Forecast of Stock Issuance Information Based on High Performance Computing and Blockchain Technology 754
Di Zhao, Yimei Cao, and Chaoying Xiao

Design and Research of Enterprise Tax Accounting Database Under the Background of Big Data 763
Yuxin Xie

Mental Disease Early Warning Analysis Technology Based on Social Media Big Data Algorithm 771
Jiangnan Li

Logistics Network Deployment Planning Under the Background of Big Data Technology 778
Meina Guo

Power Grid Parameter Acquisition System Based on the Fusion of Intelligent Unit and Edge Computing..... 787
Dawei Wang, Jianjun Liu, Bo Zhao, Xiaofeng Xu, Kenan Hou, Bing Ma, Wei Tang, Changyuan Bao, Shuwei Hou, and Chunyang Dong

Intelligent Analysis and Processing System of Financial Big Data Based on Neural Network Algorithm 795
Wei Shi

Analysis of Financial Sharing Model Based on Block Chain Technology in Internet + 5G Environment..... 802
Yu Li

Performance Prediction Model Based on K-Means Clustering Algorithm 809
Wei Zheng, Lihong Hao, and Tao Liu

The Design of Contemporary Urban Literature System Based on Internet of Things Data Processing Technology 817
Qiong Zhong

Laboratory Information System Under Big Data Technology 826
Xianjin Lin and Ziyu Zheng

Machine Learning, Computer Vision, Image Processing, and Their Intelligence Applications

Three-Dimensional Tree Visualization of Computer Image Data Based on Louvain Algorithm 837
Chao Mao

Sales Forecast for New Energy Vehicles Based on Grey Relevance Analysis and a BP Neural Network 846
Jing Jiao

Formative Evaluation of College English Autonomous Learning Based on Whale Optimization Algorithm 853
Xueqian Chi

Student Performance Prediction and Learning Intervention System Model Based on Machine Learning 862
Yuqiang Kong

Construction of Models for Evaluation and Development of Urban Tourism Resources Based on Artificial Intelligence 871
Chunling Liu

Image De-rain Algorithm in the Architecture of the Big Data Platform of the Internet of Things 879
Yongjun Qi, Haiyan Wu, and Chadi Altrjman

Intelligent Analysis Technology of Sports Training Posture Based on Deep Learning 887
Cong Zhou

Design and Implementation of English Speech Scoring Data System Based on Neural Network Algorithm 896
Xia Sun

Financial Risk Forecast Model of Small and Medium-Sized Enterprises Based on Neural Network Algorithm 905
Yu Tao

The Analysis of COVID-19 Propagation Process Based on Population Flow and Cellular Automaton Model 913
Congxiao Zhang

Facial Image Restoration Algorithm of Dazu Rock Carvings Based on Machine Vision 922
Zhuying Ran

Data Mining Model of English Translation Teaching Based on Virtual Reality Technology 930
Na Li

Prediction Algorithm of Music Trend Based on Neural Network 938
Xuyan Wang

The Application of Panoramic Image Mosaic Technology in Modern Digital Painting 946
Tao Zhang

Soccer Match Target Tracking System Based on Particle Swarm Optimization 954
Zerong Jiang and Chen Zhu

Phase Study of CT Respiratory Slice Constructed by Three-Dimensional Image Fusion Technology 963
Hanxu Zhang

Dynamic Image Monitoring Method of Hitting Point for Tennis Players in Serving Instant 971
Xiaopeng Ji and Chadi Altrjman

Cardiovascular Monitoring System Design Based on Medical Imaging Technology and Artificial Intelligence Algorithm 978
Ruiliang Ling and Amar Jain

Electrical Control System Based on Machine Algorithm and Adaptive Fuzzy Algorithm 987
Lu Zhou and Yu Cui

Intelligent Optimization Algorithm in Virtual Design of Landscape VR 995
Danping Yi and Quannan Wang

Processing Method and System of Artificial Intelligence Application Based on Open Source Deep Learning Framework 1003
Yihan Tu

Application of Computer Virtual Technology in Physical Training Video 1011
Taowen Wei

Machine Automatic Translation Quality Evaluation Model Based on Recurrent Neural Network Algorithm 1019
Beibei Ren

AHP-Bayesian Network Based Analytical Model for HRMIS 1027
Zijuan He

Medical CT Image Enhancement System Based on Swarm Intelligence Optimization Algorithm 1035
Li Xu, Yuqing Hou, Fengjun Zhao, and Jinniu Bai

Design of English Multimedia Intelligent Translation System Based on Neural Network Algorithm 1043
Zheng Zhao

Design of Glasses Products Based on Artificial Intelligence 1051
Han Sun and KieSu Kim

Author Index 1059

**Cyber Intelligence and Analytics
from Different Layers of Data
Warehouse Solutions**



Intelligent Archives Warehouse Management Based on New Information Technology

Mengxue Zhang^{1(✉)} and Rasha Almajed²

¹ Shandong Vocational College of Science and Technology, Weifang 261053, Shandong, China

² American University in the Emirates, Dubai, UAE

Abstract. How to manage the information in the modern archives of information technology sweeping the world has become an urgent issue. The purpose of this article is to study the management of smart archives based on new information technology. This article establishes a management system for smart archives, and compares it with references and cases and data obtained from the Internet. Determine the specific system functions and management process. First, it analyzes the historical background and concept of smart archives in detail, and then according to the current new information technology, taking into account the problems of drainage, dust prevention, light protection, ventilation, fire prevention, and anti-theft, it adds intelligence to the smart archive management system. Sensors and automation functions to ensure the security of physical files and electronic files. On the basis of information technology, algorithms and practices are added, and actual data is used to prove the feasibility and rationality of the research direction of this article. Experimental research shows that based on the research of smart archives warehouse management based on the new information technology proposed in this paper, the ACA algorithm is integrated into the smart archive system, and the efficiency will be greatly improved after the ACA algorithm is optimized.

Keywords: Archive management · Information technology · Archive management · Archive informationization

1 Introduction

With Internet technology, mobile communication network, Internet of Things, cloud computing, big data and other emerging information technologies have been widely used, dragging the development of file management informatization, how to establish a modern informatization archive library to protect archive information the security has become a problem for managers. First, conduct a demand analysis of archive management, design the specific operation process of the archives, and then use new information technology to realize digitization.

In the research on the management of smart archives and warehouses based on new information technology, many scholars have conducted research on it and achieved good results. The development trend of technology is still strong. People are beginning to enter the era of intelligence. With the ubiquity of technology, historical changes have taken

place in the field of file management, and the new intelligence of file management is about to open [1]. Chen Xiaolei and Wu Wei said that the development of integrated management of file information has become a trend. Facing the management of massive information resources, archivists need to explore the transformation and development of archival management. Under the Internet + mode, archives can be intelligent, modernization of information, standardization of function management, development opportunities for archives management, and efforts to build and promote intelligent archives are important guarantees for archives development [2]. The intelligentization of archive warehouses can avoid many shortcomings of traditional warehouses and achieve the improvement of warehouse management goals.

This paper determines that the basic ACA algorithm and the optimized ACA algorithm are added to the management system of the smart archives, and then the optimized ACA algorithm is compared with the heuristic algorithm, and the actual data is used to prove that this is added to the archive management system.

2 Overview of Smart Archive

2.1 The Background of the Archive

The archives are produced with the appearance of words, and the archive warehouse is the companion of the archives. In history, every dynasty will have its own archive storage institution. According to historical research, the generation of archives can be traced back to the Yin and Shang dynasties, when the society at that time would already use the old archives in the basement of the ancestral temple. The Tianfu in the Zhou Dynasty is the earliest archival storage institution that still has documents that can be studied.

Since the emergence of warehouses for storing archives, people began to pay attention to the safety of archives. From site selection to construction to use, all factors such as archive safety and long-term preservation were taken into consideration. The construction of archives in ancient times needs to fully consider the problems of drainage, dust-proof, light-proof, ventilation, fire prevention, anti-theft, etc., and there are more stringent requirements for the configuration of archives management personnel [3].

2.2 The Concept of Smart Archives

For example, some scholars believe that the definition of smart archives is the use of emerging technologies such as the Internet of Things and cloud computing to intelligently manage diversified archives. With the intelligent perception and processing capabilities of file information, it can provide ubiquitous file information similar services. Another researcher believes that starting from the idea of creating intelligent archives, he proposed that intelligent archives are a part of the city, and the way of embedding archives is combined with intelligent buildings and intelligent technology to realize intelligent management of archive entities. And the information of the city during operation [4].

2.3 Functional Analysis of Smart Archives

The purpose of building a smart archive is to do a good job in the scientific management of the archive, to discover, warn and resolve potential abnormalities and safety hazards in a timely manner, and to ensure that the loss of archive entities is caused by external factors. Reduce to a minimum. Therefore, when designing an intelligent file management system, the following key characteristics should be considered.

(1) Intelligent collection function

Intelligent management must be created on the basis of a full understanding of warehouse information. Without information, it cannot be managed. Only by submitting enough information can we effectively analyze the information and make decisions. The data collection function is the main function to be realized in the design of the intelligent file management system. The distributed smart sensor mainly collects environmental information of the file repository.

(2) Intelligent analysis function

The data collected by the smart sensor is stored in the database and compared with the pre-defined range of environmental standards in the archives to analyze and judge the hazard nature, risk factors and risk levels of the incoming data. If the incoming data is within the specified standard range, the system will automatically record the data and analyze its changing law.

(3) Intelligent control function

The intelligent control function includes automatic alarm function and intelligent connection control function. The intelligent management system compares and analyzes the collected data and information, and transmits the analysis results to the intelligent control system. The intelligent control system calls related equipment to the intelligent management system according to a predefined control program and sends out signals wirelessly. Send it to the file manager [5].

(4) Autonomous learning function

In addition to tracking and monitoring the corresponding information, the intelligent management system also restores this information to the database, and continuously adjusts and monitors behavior until the information is stable. Another function is automation. The system will automatically record every adjustment and response measures, summarize and classify.

2.4 Construction of Smart Archives

The construction of a user-oriented smart archive management system is mainly divided into: service body, perception layer, data layer, guarantee system, resource integration

processing layer, service layer, user and other parts. Under the premise of file informatization, we will continue to use new Technology, new mechanisms, and new means fully automate, intelligentize, and humanize the smart archive library.

(1) Service subject

The establishment of a smart archive database must cooperate with a government department or a certain unit. The service body includes not only professional archive staff, but also information workers from the government, enterprises and institutions [6].

(2) User

Users refer to government departments, enterprises, institutions, and the public that have certain file requirements. Provide timely, accurate, and comprehensive archive information services for information users; a smart archive library is needed to provide them with comprehensive, authoritative, and true information.

(3) Data layer

The function of the data layer is to classify and store the data that can be sensed by the perception layer. It is the foundation of the information service of the smart archives. The data mainly comes from: digital achievements of archives, intelligent perception of archived information of party and government agencies and public institutions, management process records of intelligent archives, network information collection, and intelligent collection of public archives utilization data.

(4) Resource integration processing layer

The resource integration processing layer is mainly responsible for collecting and integrating allocated resources into various databases at the data level, conducting data mining, network-based and content-based knowledge bases, providing resource integration and service environment integration. The processing process includes analysis, organization and development, Store and share.

(5) Service layer

Service is the most important link in intelligent file management. The intelligent archive management database will directly face users and provide users with information services that meet their needs through the service platform.

(6) Perception layer

Perception level is the basis for constructing the entire system and the key to realize the association of files and information. The perception and collection of relevant data are realized through the Internet of Things, and the collected data is counted, analyzed,

sorted and processed, and then the processed data is stored in the database to prepare for file sharing and services [7].

3 Design and Implementation of Smart File Management System

3.1 System Development Platform and Functional Design

The development of a smart library management system should take application as the core, simplicity, use, and convenience as the base point, and take into account technology, application and development.

The development of the smart file management system uses the .NET technology, the language uses the C# language, and the SQL SERVER is selected as the system database on the C/S-based system architecture.

Advanced systems must ensure that the interface is simple, easy to use, and fast. In terms of functions, it must (1) provide forward and reverse functions to ensure that users can switch to another function at any time. (2) The home page return function allows the operator to directly return to the main interface at any position. (3) The refresh function allows the operator to refresh the system data at any time. (4) Timing reminder interval can be flexibly set. (5) The current task should be reminded. (6) You can switch from the task reminder window to the corresponding task window [8].

3.2 Intelligent Structure of Archives

Archives are a permanent base for keeping archives. Only under certain temperature and humidity conditions can archives be preserved for a long time, which can well avoid mildew and insects. Therefore, it should be clear that the temperature and humidity of the archives should be controlled within a certain range. In order to prevent thieves from stealing, destroying the archives of the warehouse, and tracking accidents in time, real-time video surveillance is designed [9].

3.3 Application of ACA Algorithm

In the information flow of the process of retrieving archive information keywords, the ACA algorithm can be used to concretely represent the retrieval process. The variable τ_{ij} represents the amount of information from i to j on the channel:

$$\Gamma = \{\tau_{ij}(t) | c_i, c_j\} \quad (1)$$

Among them, set C is a collection of residual information of elements. Assuming that the content of path pheromone at the beginning is $\Delta\tau_{ij}(0) = 0$, then the transition probability from channel i to channel j from k to time t can be obtained as $p_{ij}^k(t)$:

$$p_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha [\zeta_{ik}(t)]^\beta}{\sum_{s \in \text{allowed}} [\tau_{is}(t)]^\alpha [\zeta_{is}(t)]^\beta} \quad (2)$$

Among them, the heuristic factor of the expression formula, the larger the probability of k passing through the information channel [10].

4 Data Practice Based on ACA Algorithm

4.1 ACA Algorithm Comparison

As shown in Table 1, the ACA algorithm is compared with the traditional heuristic algorithm, and the efficiency of the algorithm search is obtained by classifying and searching the files in the archives in the same world. Until 20 s ago, the number of archives retrieved by heuristics was greater than the number of archives retrieved by ACA algorithm. For example, 0.15 more files were retrieved in 10 s.

Table 1. Basic ACA algorithm test data

Algorithm name	5S	10S	15S	20S	25S	30S
ACA	0.20	0.75	1.3	3.0	3.6	5.0
Heuristic	0.40	0.90	1.4	2.95	3.5	4.30

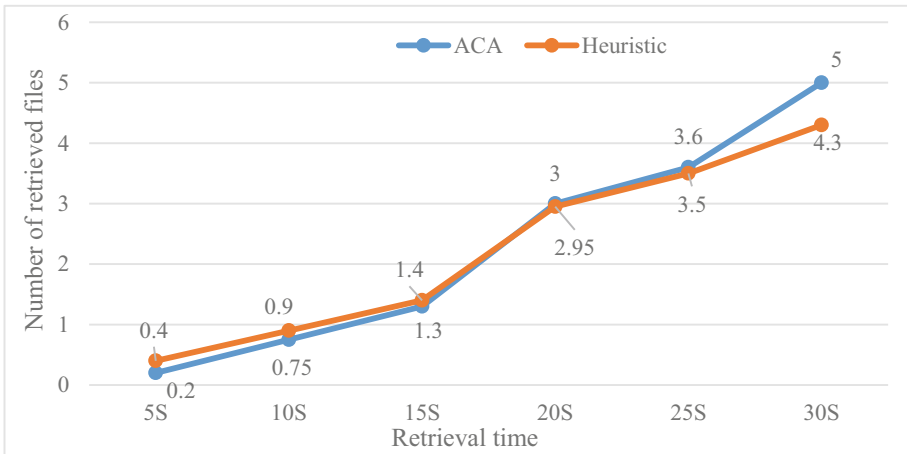


Fig. 1. Comparison of algorithm retrieval efficiency

As shown in Fig. 1, the heuristic algorithm is more efficient in the first 25 s of retrieval, but after 25 s, the retrieval efficiency of ACA has begun to continuously improve and has far surpassed the heuristic algorithm. So we come to the conclusion that there are generally more content to be retrieved in the management system of the intelligent archives, so it is more appropriate to use the ACA algorithm, because the ACA algorithm is more suitable for retrieval with a large amount of archive data [11, 12].

4.2 ACA Algorithm Optimization Data Comparison

On the basis of the preparation of 120,000 archive data, the heuristic algorithm, basic ACA algorithm, and optimized ACA algorithm were searched for the same data on the same computer.

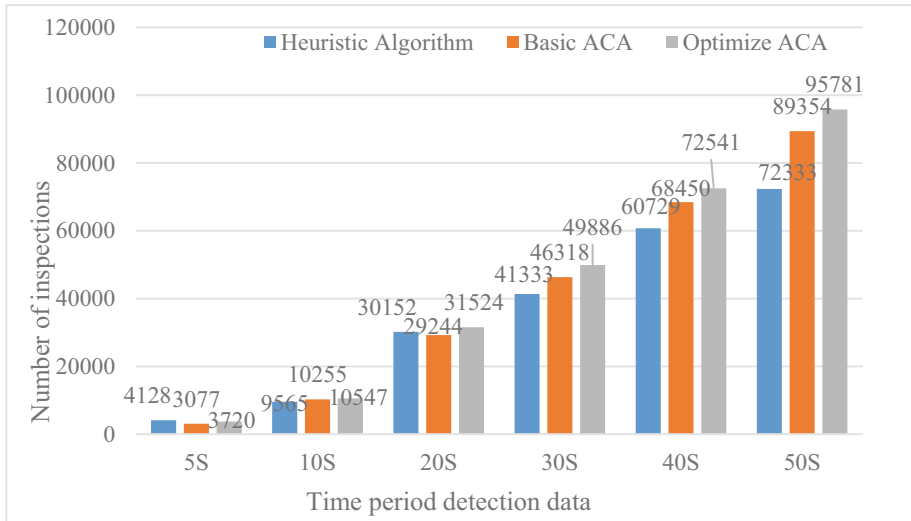


Fig. 2. Comparison of algorithm retrieval data

As shown in Fig. 2, the actual test shows that the heuristic algorithm still occupies a certain advantage in the first 20 s, and is more efficient than the basic ACA algorithm, but after the ACA algorithm is optimized, it is faster than the first two algorithms, and in 20 s In the future, the retrieval efficiency will far exceed other algorithms.

5 Conclusions

This article uses data obtained from references and the Internet to set the algorithm of the intelligent archive management system, and adds intelligent system functions on the basis of planning the system operation process to ensure that users and administrators can carry out the corresponding operate. And on the basis of optimizing the ACA algorithm, a specific practical analysis of the efficiency of file detection is carried out, and the corresponding conclusions are obtained and compared, which proves the superiority of the algorithm over other efficiencies.

References

1. Yang, S.: Talking about promoting the construction of smart archives. *Inside Outside Lantai* **303**(30), 29–30 (2020)
2. Chen, X., Wu, W.: The future of archives in the era of “Internet+”——Smart Archives. (2020–5), 178–179 (2021)
3. Urakseev, M.A.: Intelligent information-measuring systems with sensors based on diffraction grating. *Chem. Technol. Control Manage.* **2018**(3), 11–12 (2018)
4. Hamdy, W., Mostafa, N., Alawady, H.: An intelligent warehouse management system using the internet of things. *Int. J. Eng. Technol. Sci.* **32**(1), 59–65 (2020)

5. Mohamed, M.: Smart warehouse management using hybrid architecture of neural network with barcode reader 1D/2D vision technology. *Int. J. Intell. Syst. Appl.* **11**(11), 16–24 (2019)
6. Zou, T., Bai, S.J., Rong, X.B.: Research on marketing management system based on independent ERP and business BI using fuzzy TOPSIS. *J. Intell. Fuzzy Syst.* **40**(4), 1–9 (2021)
7. Vijay Kumar, T.V., Arun, B.: Materialized view selection using HBMO. *Int. J. Syst. Assur. Eng. Manage.* **8**(1), 379–392 (2015)
8. Lu, P., Liu, P., Xu, J.: Design of intelligent warehouse management system based on MVC. *Int. J. Adv. Network Monit. Controls* **6**(2), 79–87 (2021)
9. Mohamed, S., Alshalfan, K.A.: Intelligent traffic management system based on the internet of vehicles (IoV). *J. Adv. Transp.* **2021**(4), 1–23 (2021)
10. Chen, Z., Chen, R., Chen, S.: Intelligent management information system of urban planning based on GIS. *J. Intell. Fuzzy Syst.* **40**(1–2), 1–10 (2020)
11. Li, Y., Shi, X., Diao, H., et al.: Optimization of warehouse management based on artificial intelligence technology. *J. Intell. Fuzzy Syst.* **6**, 1–8 (2021)
12. Kashevarova, G., Semina, A., Maksimova, S.: Intelligent and digital technologies in the construction objects technical diagnostics. *Int. J. Comput. Civil Struct. Eng.* **17**(2), 22–33 (2021)



Design and Algorithm of File Management System Based on RFID

Weimeng Liu^(✉)

Financial Department, Shandong Polytechnic College, Jining, Shandong, China

Abstract. Document management is the original documents of different forms with storage value directly produced by the subject in various activities, which directly record the key information of the subject, which has great social practical value. The traditional method of document management is not only time-consuming and laborious, but also easy to make mistakes. Document management is a lot more difficult. These backward management methods can no longer meet the actual needs of current economic and social development. However, with the advent of the Internet age and the popularization of computers, the use of computer technology to develop effective digital document management can greatly improve the company's operating efficiency, thereby enhancing the strength of the entire enterprise. This article focuses on the research of RFID-based archives management, understands the related theories of archives management and RFID based on the literature, and then designs the RFID-based archives management system, and then tests the designed system. The test results are obtained. It is concluded that the system designed in this paper can meet the design requirements, and the recognition rate is not much different from the actual value.

Keywords: File management · RFID technology · System design · Electronic information

1 Introduction

With the vigorous development of network information technology, the network has also begun to infiltrate all areas of daily life, including the transformation of traditional office work methods [1, 2]. Using smart office methods to replace traditional manual methods, greatly improving office work efficiency [3, 4]. Because the traditional document management method is relatively backward, time-consuming, resource-intensive, and it is more difficult to maintain the document in the later stage. Therefore, with the development of modern computer technology, the use of modern computer technology to achieve document management has become the main method of document management [5, 6]. Modern document management uses the network and computer as the main media, and takes the new document management theory as the main technical guiding principle of file management to scan, store and use documents [7, 8].

For archive management research, some researchers suggest that the file management system must first use a logical storage structure to manage these files. Because

today's companies have to deal with more and more complex files, business-level file management systems need to organize and design a reasonable number of files. Documents, and use advanced network systems to quickly and accurately provide users with search results [9]. Some researchers pointed out that defining user permissions is also a very important aspect. Since the file management system includes all parts of the entire enterprise, and the file operation authority of each department is different, it is necessary to logically set the user authority of the file management system [10]. Some researchers also pointed out that traditional file management requires the file manager to manually search for paper files. Online file management significantly reduces file search time and improves the use of corporate file value. In the storage of records, operators can quickly and easily find the files they need, which also improves the work efficiency of enterprise employees [11]. To sum up, although the file management system has been researched, its application is not very extensive. There are still many companies in China that have not realized the informatization of file management. This shows that there are still some problems in the file relationship system at this stage.

This paper conducts research on RFID-based archives management, analyzes the requirements of archives management system and the feasibility of RFID technology on the basis of literature data, and then designs and tests the designed system based on RFID, then draw related theories through test results.

2 Research on Archives Management System

2.1 File Management System Requirements

Archives are valuable historical archives formed on the subject of various social activities. These subjects can be government agencies, social organizations or individuals. "Archives are the common intellectual property rights of mankind. With the development of countries, enterprises, and institutions, the number of various information files is increasing day by day, all kinds of printed information are emerging in large numbers, and the number of files lacking management is gradually increasing. Traditional file management can no longer meet actual needs. The details are as follows.

- (1) The document workflow is very complicated and inefficient, and often takes a long time. File management is done through traditional manual editing, which is cumbersome and inefficient.
- (2) The file storage sequence is easily interrupted. Saving files is confusing, it is difficult to find files, and manual operations are unavoidable to make mistakes.
- (3) It takes time to search for files. The process of doubling the file size requires filtering all files to find a specific file.
- (4) File security is weak and cannot effectively protect file security and prevent file damage from malicious theft.
- (5) The amount of file information is small. Traditional management methods cannot manage file management information, and manual recording makes it difficult to timely and accurately count file usage.
- (6) The efficiency of file inventory is low. Due to the large number of files, it is difficult to imagine the workload of counting all the files, which can be a huge task.

2.2 The Feasibility of RFID Archive Pipeline

(1) Economic feasibility

The cost of using RFID file system includes three main parts: hardware cost, software cost and labor cost. Figure 1 shows the cost structure of the RFID system. The hardware cost mainly includes electronic tags, antennas, readers and other related hardware. Software costs mainly include computer, server platform and system software development costs. Labor costs refer to labor costs that are mainly invested in use and system maintenance. The total cost of each component is the cost of the entire system.

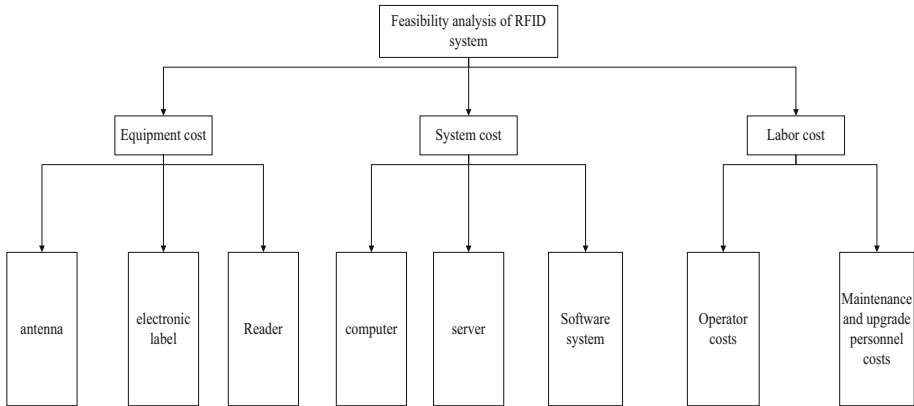


Fig. 1. Cost structure diagram of RFID system

(2) Technical feasibility analysis

Nowadays, the development of RFID technology is becoming more and more mature, and it is widely used in various system integration and data management platforms. Leading software developers have also provided related interfaces and solutions to support the RFID technology in their products, and conducted a large number of product tests and application promotion. Compared with other identification technologies, the advantages of RFID technology can solve or alleviate many problems in the file management system, so this is a very promising technology. For example, RFID technology can be used to read batch archived information, but the barcode system requires a single scan and input. RFID technology supports real-time on-site file detection, which cannot be achieved by a simple barcode system. In addition, when detecting the access control

unit, the barcode system must use magnetic stripe technology for anti-theft protection, while RFID technology can be carried out without resorting to other technologies [12].

(3) Safety feasibility analysis

The application of RFID technology to the file management system, the data transmission between the reader and the electronic tag, the data transmission between the background computer system and the reader, all have specific security risks, but this security risk is encrypted and may use computerized encrypted information to solve. Electronic storage tags have no special security settings, and there is a unique and immutable serial number in the tag. The storage area can store data and can be read and written without a security authentication mechanism. The security of this type of label is not very high, but some encryption methods can be used in the application system to ensure the security of the data. Electronic encryption logic tags are different from electronic storage tags. The tag chip has functions such as logic encryption circuit and one-time programming. After this mode is enabled, a verification key is required to read and write internal data, and one-time scheduling can prevent data from being leaked. This is more secure than electronic tags. Therefore, the use of logical encryption tags in the file management system can effectively improve the security of the system.

2.3 RFID Algorithm

The basic binary search algorithm of the RFID system uses Manchester information encoding and is executed using a series of special reader instructions according to the unique serial number of the tag.

The specific workflow of the BST algorithm is as follows:

- 1) During the execution of the system, the reader first issues a request command. The first command parameter is 11 ... 1 (the number is the same as the number in the tag ID), and subsequent adjustments are made according to the status. The sender's receiving tag compares the value of the command parameter with the value of its own ID. If her ID value is less than or equal to the value of the command parameter, the ID is sent to the reader. Otherwise it will not respond.
- 2) The reader judges whether it has received the information returned by the tag, if not, then execute step 7).
- 3) If the ID information returned by the tag is received, judge whether there is a conflict. If there is a conflict bit, reset the request command parameters, set the received maximum multitag ID to 0, and do not change the previous bit. All the following bits are set to 1, take it as a new parameter and return 1).
- 4) If the reader determines that there is no conflict, the reader sends a selection command. The parameter is the received tag ID, select the tag, and then send a data read command to the tag to read and write data.
- 5) The reader issues a cancel selection command, and the parameters match the parameters of the selection command.
- 6) The reader resets the parameters to 11 ... 1 and returns to step 1).
- 7) The program ends.

The number of request commands required by the BST algorithm to identify the first tag is

$$f(N) = \log_2 N + 1 \quad (1)$$

Where N is the number of tags.

The total number of request commands to be sent when all N tags are recognized is

$$SUM(N) = \sum_{i=1}^N f(i) \quad (2)$$

3 Design of RFID-Based Accounting File Management System

3.1 The Overall Functional Architecture of the System

System configuration management functions include sorting file management, file numbering rule management, storage plan management, metadata format management, category definition, etc. File management functions should include file collection, metadata, and log records. File security management functions include authentication, principle management, follow-up control, merging, storage, use, statistics, authentication, discarding, transmission, backup, etc. This feature plays an important role in ensuring the reliability of accounting files. System parameter management, system users and resources, system function configuration, etc., constitute system management functions.

3.2 Data Acquisition Module

The data collection and processing part is divided into five parts: file box data collection, file bag data collection, loan return data collection, safety management data collection, and borrower data collection. The data module uses RFID equipment to understand archives and personnel information, and the 485 communication module understands the specific status of the vibration sensor, as well as specific security information such as door status and files, and reports to the administrator. The data entered by the data collection unit or part of the reported data is stored in the database through the file management unit.

3.3 File Management

(1) Archive

The archiving of accounting files mainly includes archiving methods (online automatic, manual, etc.), archiving time, grade archiving, classified archiving, delivery archiving, etc.

Filing evaluation mainly refers to the division of the retention period of accounting vouchers, which can check the authenticity, reliability, completeness and usability of accounting vouchers, and can only be archived after the evaluation is passed. The system can integrate the detected features into the file interface, reducing manual intervention.

(2) Sharing

For public responsibility files, financial or file management personnel send the published accounting data to the server via the Internet, and share the data for users to consult and use. You can also send online (QQ, etc., Email) and send accounting documents for applicants' reference and use.

For confidential information, users can access the information through the application client or the Internet approved by the user. In addition, you can also directly access and copy files in the financial office. However, this method of copying may lead to unauthorized distribution of accounting information. Enterprises need to take effective control, borrow, copy, obtain, and destroy accounting files in a timely manner, and obtain and destroy copied accounting records in a timely manner.

(3) Disposal

Disposal of accounting files: identification and destruction. The company must organize accounting, auditing, archives and other departments to evaluate the expired accounting files, read them file by file, and put forward evaluation opinions. For accounting records that need to be retained, a 10-year or 30-year confidentiality period is redesigned according to the content, and accounting records with unreserved value are destroyed after the confidentiality period ends.

3.4 System Management Module

The system management module is mainly used for system administrators and file administrators to maintain the system and data. There are four main types of enterprise file management system users: system administrators, file administrators, file approvers, and system users. In this article, the system administrator performs the user authentication function. The file manager performs the task of importing and sorting group files in this article. System users are divided into advanced users and ordinary users with file control permissions. Ordinary users can use the workflow "submit application to advanced users (advanced users)". When the application is up and running, ordinary users can download files to their desktop computers to facilitate their office work.

4 System Test

After the system design is completed, verify and test the various functions of the system according to the file management procedures such as data collection and processing, file lending, and file return to ensure the normal and stable operation of the system. The test results are shown in Table 1.

Table 1. System test results

	Actual results	Test results
Portfolio inspection	150	145
File box detection	150	148
File bag and file box mixed detection	150	143

**Fig. 2.** System test results

It can be seen from Fig. 2 that the system RFID has a high recognition rate, and the recognition result is not much different from the actual result, and the recognition rate is about 94%.

5 Conclusions

This article focuses on the research of RFID-based file management. After a general understanding of the relevant theories, the RFID-based file management system is designed and the designed system is tested. From the system test results, the detection value of the system is the actual value is not much different.

References

1. Li, H., Zhang, G.: Coastal atmospheric climate and vehicle file management based on high-resolution image recognition. *Arab. J. Geosci.* **14**(11), 1–16 (2021)
2. Park, J.E., Park, Y.H.: Fog-based file sharing for secure and efficient file management in personal area network with heterogeneous wearable devices. *J. Commun. Networks* **20**(3), 279–290 (2018)

3. Otoo, F.: Human resource management (HRM) practices and organizational performance. *Empl. Relat.* **41**(5), 949–970 (2019)
4. Gupta, S., Kumar, S., Singh, S.K., Foropon, C., Chandra, C.: Role of cloud ERP on the performance of an organization: contingent resource-based view perspective. *Int. J. Logist. Manage.* **29**(2), 659–675 (2018). <https://doi.org/10.1108/IJLM-07-2017-0192>
5. Ga Mido, H.V., Ga Mido, M.V., Sison, A.M.: Developing a secured image file management system using modified AES. *Bull. Electr. Eng. Inf.* **8**(4), 1461–1467 (2019)
6. Shipton, H., Budhwar, P., Sparrow, P., et al.: Human resource management and innovative behaviour: considering interactive. *Informal Learn Act.* (2016). [https://doi.org/10.1057/9781137465191\(Chapter3\):32-47](https://doi.org/10.1057/9781137465191(Chapter3):32-47)
7. Wen, K., Chai, D., Wang, J., et al.: Discussion on risk-based standard management of trial file management in clinical trial institutions. *Chin. J. New Drugs* **27**(24), 2916–2921 (2018)
8. Zhu, X.G., Shen, Y.: MapReduce optimisation information query method for file management system. *Int. J. Reasoning-Based Intell. Syst.* **10**(2), 117–121 (2018)
9. Choukse, D., Kumar, U.: Enterprise web based file management – a system architecture and solution. *Int. J. Comput. Appl.* **179**(12), 48–51 (2018)
10. Jiang, X.: Framework research on embedded file management system of private colleges based on CAN bus. *Rev. Fac. Ing.* **32**(15), 246–252 (2017)
11. Lee, J., Nam, S., Kwak, S., Park, S.: Context-aware file I/O management system for mobile devices. *IEEE Access* **9**, 16556–16566 (2021)
12. Kumar, A., Gopal, K., Aggarwal, A.: A novel and efficient anti-collision protocol for RFID tag identification. *J. Commun. Software Syst.* **12**(4), 173–181 (2016)



Design of Information System for Accounting File Management Based on Computer Technology

Yan Sun^(✉)

Department of Accounting, Shandong Polytechnic College, Jining, Shandong, China

Abstract. The tremendous progress of scientific and technological information has greatly improved the efficiency and level of corporate accounting file management, and its positive significance is obvious to all. The purpose of this article is to study the design of the information system of accounting file management. This article studies the background of the accounting archives management information system and discusses the importance of the research. This article conducts a detailed demand analysis, including capital requirements and functional requirements. This article investigates and researches the accounting file information system of a company in this city. Experimental research shows that the company's accounting information system data information security level is generally considered to account for 45%, which is a large proportion, and the company itself lacks a set of rigorous standards in its management system. The internal control system and data encryption measures are not in place.

Keywords: Accounting file · File management · File system · Information system

1 Introduction

Nowadays, the company's file management system has the needs of large amount of information, real-time updates, and frequent access. This is a major challenge to the current accounting file management system and methods [1, 2]. In the past, accounting file management was managed by human resources. However, with the increase of file resources, file management is still taken over by human resources, which will lead to problems of low performance and increase of manual errors. If the file information import function is complete, complete and systematic, it will undoubtedly provide great flexibility to the file management function [3]. Therefore, it is particularly important to use computers to manage financial files.

In the research on the design of the information system of accounting file management, many scholars have studied it and achieved good results. For example, Kulagina ZD believes that statistical file information can be used to identify the company's assets in different periods, and the company should adopt a file management system to unify Store [4]. Onodi B E believes that the centralized storage and centralized control of all

electronic databases can be realized by adding the function of the archive unit to the information system [5]. It can be seen that more and more companies are implementing electronic accounting file management.

This article studies the accounting file management system under the conditions of informatization. This article analyzes the problems that need to be solved in the electronic management of accounting files under the conditions of informatization, and puts forward suggestions and suggestions. Finally, take a company in this city as an example, apply the preliminary decision to the company's file management system, and record the results of online file management.

2 Research on the Design of the Information System of Accounting File Management

2.1 System Function Module Design

(1) Functional structure design

System management system functions include command management, user control, role control, backup and data recovery, system configuration and other functions.

User information can be added and modified in the user control. Existing users can define roles. If an existing user has a role, then when the user is deleted or deactivated, all corresponding roles corresponding to the user will be deleted. In execution control, you can share existing commands, configure administrator relationships, and select programs or services that can be used [6, 7]. You can add management roles, delete roles, and adjust the relationship between roles and permissions. If a role is deleted, the unit corresponding to the unit will be deleted together. In addition, you can back up and restore system data and set the life of the system.

(2) Physical management module

In physical file management, physical file files can be managed, including storing physical files, exporting physical files, managing physical files in credit unions, returning physical files, saving new or additional files. The library must sort the physical data it generates according to specific specifications, and the system shares the corresponding storage location to record this file. The delivery of physical files refers to the completion of the physical processing of data. Sometimes the final file may be damaged or moved to the storage location [8, 9]. When downloading a physical file, the program will check whether the user has a rental order. If it is in the rental order, it will record user information

(such as user name, rental time, return time, etc.) and mark the file as borrowed. When the user returns, the broken mark of the recording file is cancelled and the recording is performed. Physical files can be added at any time. If there are new physical files to be stored in the storage, they must be added.

(3) Electronic file management module

The electronic document management system includes electronic document registration control, electronic document conversion and deletion, electronic document request and electronic document statistics. The central electronic file system is created by a unified electronic file and image file interface. Electronic files include file creation, editing, addition, modification, deletion, request and management services.

- 1) Electronic file import control. The files to be downloaded are printed and stored by the machine, and the files and procedures must be checked by the administrator before converting and deleting electronic files.
- 2) Electronic file statistics provide statistical analysis of specific users, specific transactions, and business data at specific times. Users can customize the title, query location, data format and other information. Display data and analysis results on the page. Statistics is an ideal choice based on statistics, design statistics and statistics based on average conditions.

2.2 Electronic File Management

This function plays a very important role in the stability and quality of digital information. All file pictures have stricter procedures.

(1) Business information map

This part is the most important part of the preparation of archive library resources, and it is also the basis of the entire program. This part is to complete the construction of the archive art library. The archived archive images are sorted by type and exported to the server. The server declares and saves the captured archive images, and creates an archive image file library [10]. When browsing archive image files, a catalog file is created according to the archive image data information, and the archive image is downloaded and displayed according to the catalog.

- (2) Establishment of structured information
 - 1) Schedule information file, while creating an image file library, integrate file data information into the database for easy retrieval and reading when needed. That is data + image information. Obtaining two locations not only ensures the reliability and reliability of stored information, but also manages file management.
 - 2) Check the configuration information and compare whether the data and image information are compatible. As a supervisory role in processing structured information.

- 3) Structural information request, which is the query and security of large storage information. Since query and security are taboo, data requirements and security must be balanced. Therefore, it is necessary to separate the server, query the program and write the service plan. Such a design can be easily adapted to actual conditions.

2.3 Electronic File Management Design

(1) Maintenance of electronic file information

The functional steps of file maintenance are as follows.

Step 1: Log in to your account and log in to the corresponding menu view. The purpose of this step is to separate permissions so that different personnel can see different program interfaces. When the connection fails, the server saves the user information in time. Easy to upload images and archive file storage [11, 12].

Step 2: The administrator selects the file maintenance view type according to the file type. Different files have different characteristics and requirements, it is easier to maintain and request and speed up requirements of different authorized personnel.

Step 3: Capture the image. After the administrator uploads the pictures to the server, the system sorts and saves them according to different file types and file dates, which is convenient for maintenance personnel to quickly locate the files in the image library. If you know that the image is in the server folder (not the system file system), you can automatically copy it to the specified folder to complete the file customization process.

Step 4: Add information in the first electronic file information panel, and then create an image list in the sub-table. The purpose of this is to create a master table for the document for easy retrieval and reading.

Step 5: Leave the form information. The main purpose is to submit the form to the electronic file viewer. The second purpose is to reduce the data of the maintenance interface and facilitate the work of the nursing staff.

(2) Client design for file review

The future goal of document review is to make some changes to the design of the maintenance unit in order to preserve the original design and reduce the cost of maintaining and developing the document system. Under the maintenance unit, the functions of file creation, storage, and retention have been cancelled, and the functions of data sharing and file status modification have been added. That is, delete storage key, upload, add password, registration, shared data, shared file key, and change the read display mode to read-only mode.

2.4 Risk Analysis and Evaluation of Accounting File Information System

The risk level quantification table can be obtained, taking the calculation process of X1 as an example; assuming that the quantification value of the risk level of risk factor X1 is R_x , the risk probability quantification table of the accounting archives information system construction project and the accounting archives information system construction project

risk The impact quantification table shows that the quantitative value of risk impact $RA = 2,7$, and the probability of risk occurrence $RB = 0.25$, the formula is as follows:

$$RX = RX1 + \frac{(RA - RA1)(RB - RB1)}{(RA2 - RA1)(RB2 - RB1)}(RX2 - RX1) \quad (1)$$

Then by using the BORDA ordinal value method, first calculate the BORDA ordinal number of each identified risk factor, according to the ordinal value calculation formula:

$$B_X = \sum_{i=1}^2 (N - N_{xi}) \quad (2)$$

And by calculating the average value of the scoring data in the risk impact quantification table and the risk probability quantification table.

3 Experimental Investigation and Research on the Design of Accounting File Management Information System

3.1 Interview Survey

In order to systematically and comprehensively investigate the implementation status of a company's accounting file information system, the questionnaire survey method is not suitable for collecting information and questions through interviews and surveys to make targeted supplements, so as to combine the actual situation of various aspects to propose scientific research for the subject. Strategy, this survey is mainly aimed at the financial management personnel (financial section chief, accountant) of a certain company in this city.

3.2 Sources of Experimental Data

This questionnaire survey mainly adopts the form of on-the-spot questionnaires. The time span from design to issuance is from May to November 2020, which takes more than six months. In order to make the statistical conclusions more pertinent, all employees of the Finance Department will be supplemented in the follow-up. A total of 300 questionnaires were issued for this survey, and a total of 280 were collected. At the same time, the collected questionnaires were checked and screened. Incomplete and highly uniform and extreme questionnaires were excluded. In the end, a total of 265 valid questionnaires were successfully collected. Because the sample size of the survey data in this article is not complicated, we mainly use EXCEL statistical analysis software to summarize and analyze the original data of the questionnaire survey responses. The statistical data analysis in this article mainly uses descriptive statistical analysis methods.

4 Investigation and Analysis of the Design Experiment of Accounting File Management Informatization System

4.1 Irregular Internal Control and File Management

Lack of systematic internal control standards. From the field investigation and questionnaire survey, we learned that the company’s current accounting information system internal control system is not perfect and strict, and the confidentiality and security level of data information is low. The experimental results are shown in Table 1.

Table 1. Analysis of accounting Wen'DANG information control management

Confidentiality	Male (%)	Female (%)
Very safe	5	3
Relatively safe	6	4
Safer	15	17
Generally	42	48
A bit safe	11	9
Relatively unsafe	12	16
Not safe	9	3

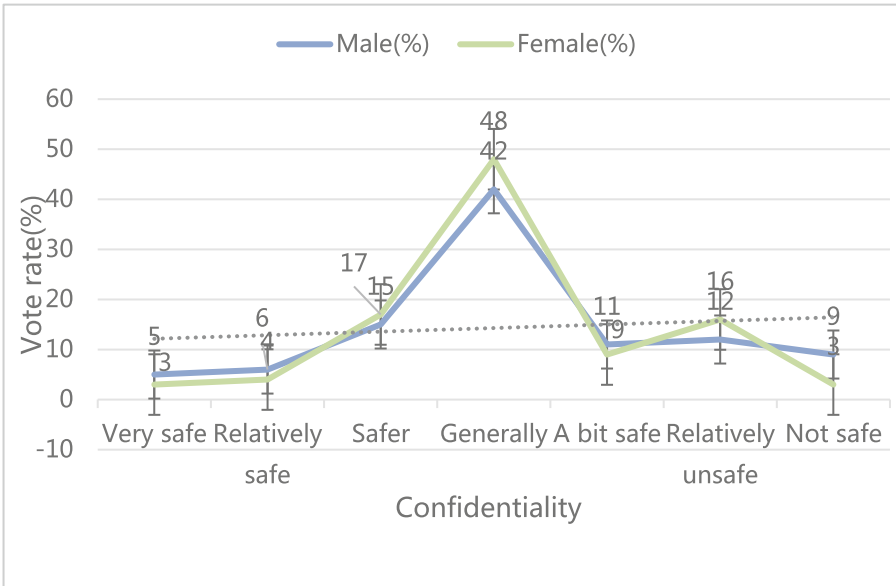


Fig. 1. Analysis of the information control management of accounting documents

As shown in Fig. 1, from the analysis of the survey data, it can be seen that the overall data of the company's statistical information system is considered to be 45% of the security level, which is relatively large, and the system management link is weak. The security and confidentiality of data information is also dangerous and can be easily destroyed. For other problems and safety issues, the main reason for the poor management of the company's internal accounting system is that the current internal control system and the corporate accounting system do not match the pace of development over time. Except for the lack of control, the company itself does not have a strict system. Without data encryption measures, scripts and computers are not properly integrated to monitor each cycle, resulting in a failure mode of operating internal controls.

4.2 Analysis of the Recognition of Accounting Staff's Status in File Management

The file management work lacks strict management standards and professional teams, and the management's positioning of file management work is unreasonable. The experimental results are shown in Table 2.

Table 2. Cognition analysis of employees's management work status

Understanding	Male (%)	Female (%)
Very important	16	20
Relatively important	17	19
Generally	44	36
Relatively unimportant	14	12
Unimportant	9	13

The results of the experiment are shown in Fig. 2. From the survey and statistical analysis, the base management and staff believe that the total file management status is 40%, which is a considerable amount. The company's current file management infrastructure is convenient, there is no full-time electronic file management personnel, there is no weight management and security of the online resource library, and many downloaded files are easily damaged and lost. There are many reasons for the inefficiency of electronic file management, except for investment in this industry. Institutional management, professional file management team, etc. As a result, the current company's electronic file management cannot be effectively managed.

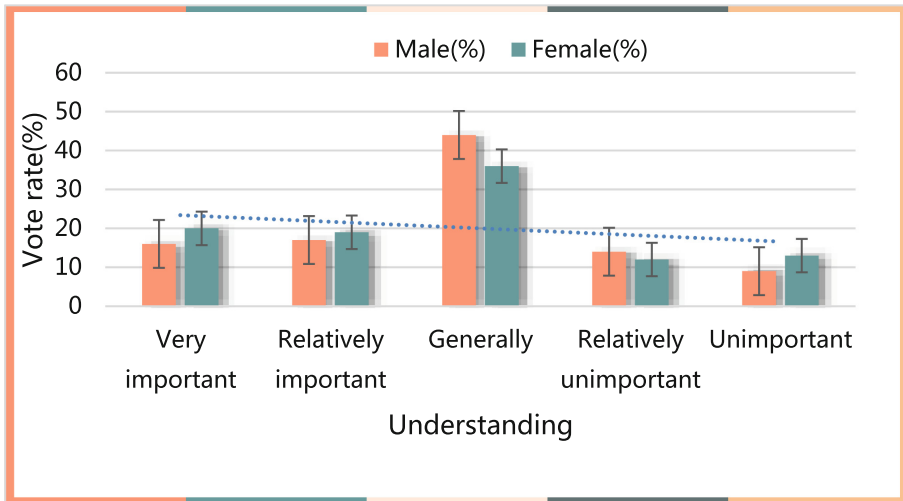


Fig. 2. Cognition analysis of employee on archival management work status

5 Conclusions

This article starts with the basics and methods of electronic file management, and summarizes the basic methods of electronic file management and the preparatory work required. Among them, the basic approach is one is the rationalization of statistical files, and the other is the docking of business plans with the electronic statistical file management system to achieve centralized and unified management and complete data management. Among them, the preliminary preparation requires the establishment of a statistical information and electronic statistical archive management system, the establishment of professional talents, and the development of a statistical archive management system.

References

1. Ruggeri, D., Rizza, C.: Accounting information system innovation in interfirm relationships. *J. Manage. Control.* **28**(2), 203–225 (2017)
2. Chen, P.-C., et al.: An information system for sustainable materials management with material flow accounting and waste input–output analysis. *Sustainable Environ. Res.* **27**(3), 135–145 (2017)
3. Kou, A.: Pre-management? How to best prepare: a brief overview for potential accounting, finance, and information systems students. *Marriott Stud. Rev.* **1**(3), 4 (2017)
4. Kulagina, Z.D., Kuchukova, N.M., Mukhamadieva, E.F.: Improving accounting, tax and management information systems as a factor in ensuring the economic security of an organization. *Bull. USPTU Sci. Educ. Econ. Ser. Econ.* **3**(33), 66–72 (2020)
5. Onodi, B.E., Ibiam, O., Akujor, J.C.: Management accounting information system and the financial performance of consumer goods firms in Nigeria. *Eur. J. Bus. Manage. Res.* **6**(1), 112–120 (2021)
6. Miftah, D., Julina, J.: Does innovation affect company performance? exploring the mediation effects of management accounting information systems. *Int. Res. J. Bus. Stud.* **13**(2), 189–200 (2020)

7. Potyshnyak, O., Litvinov, A., Danylenko, V.: Management accounting information system as a component of the global business space: the experience of Eastern European countries. *Ukr. J. Appl. Econ.* **4**(3), 57–66 (2019)
8. Zhao, Z.: 90 framework and development trend of management accounting information system—based on the empirical analysis of guangju energy chemical listed companies. *bol. Tec./Tech. Bull.* **55**(9), 701–708 (2017)
9. Rachmawati, R., Saudi, M.: Environmental uncertainty, quality of management accounting information systems and quality of management accounting information. *J. Adv. Res. Dyn. Control Syst.* **11**(3), 981–986 (2019)
10. Darma, J., Susanto, A., Mulyani, S., et al.: The role of top management support in the quality of financial accounting information systems. *J. Appl. Econ. Sci.* **13**(4), 1009–1020 (2018)
11. Aljarallah, M.K.: Accounting information systems and knowledge management. *Int. J. Managing Public Sect. Inf. Commun. Technol.* **11**(3), 1–9 (2020)
12. Afrizon, A.: Influence of top management support on quality of accounting information systems at insurance companies, evidence in Indonesia. *J. Eng. Appl. Sci.* **14**(8), 2532–2537 (2019)



Substation Cable Laying Schedule Based on GIM 3D Model

Hailiang Wu^(✉), Lin Xing, Nan Cheng, Mingfu Li, and Jizhi Su

Economic and Technological Research Institute, State Grid Hebei Electric Power Co., Ltd.,
Shijiazhuang, China

Abstract. Research on refined construction management based on three-dimensional design technology is the need to explore the construction and management of power transmission and transformation projects in the new era. As the size of power grid construction and the scale of individual projects become larger and larger, the difficulty of project construction site management becomes more and more complicated, and the project data becomes more and more complex, which puts forward higher requirements on the management level of power transmission and transformation construction projects. Traditional methods the construction management of China has been unable to meet the management needs of large-scale power grid engineering construction projects. At present, the construction of power grids involves many subjects, and the management work of design and construction, construction and completion acceptance is not effectively connected, and there are problems such as limited information interaction and insufficient information sharing. At present, the industrialization of construction is also advancing continuously, and new requirements are put forward for the process management of the project. There are new challenges in information transmission, data decision-making, etc. Only the refined management of construction based on GIM 3D design technology can meet the needs of industrialization of construction. The purpose of this paper is to study the cable laying schedule of substations based on the GIM three-dimensional model. This paper uses the GIM three-dimensional model to design the cable laying of the substation. First, design the layout of the substation, and then monitor the progress of the substation cable laying. It is concluded that the method of substation cable laying based on the GIM three-dimensional model is beneficial to the adjustment and adjustment of the schedule. Optimization has greatly improved the progress management process.

Keywords: GIM · Three-dimensional model · Cable laying · Substation construction

1 Introduction

With the rapid development of electronic technology, information technology, and network communication technology, the cable laying of substations is shifting towards three-dimensional digitalization with data and information models as the core [1, 2].

The digitization of design objects is the most basic work, especially the versatility of design object models and data exchangeability are in line with the important requirements of power grid companies in the digital handover standard of power transmission and transformation projects [3, 4]. Therefore, it is necessary to conduct in-depth analysis and research on the model range, depth, standardization requirements and application prospects of the three-dimensional digital results of substation engineering from the perspective of digital engineering construction [5, 6].

Due to the early start of the power industry in developed countries such as Europe and the United States, after a long period of development, its power pattern has been basically formed, and the power demand is relatively stable [7]. At the same time, affected by the current global economic downturn, the number of foreign power engineering projects has grown slowly, and the number of new substation projects is relatively small. In recent years, the application of BIM technology in the construction industry has greatly improved the efficiency and quality of construction projects. The vigorous promotion and wide application of BIM technology in the construction industry has also attracted great attention from foreign power construction units [8]. Due to the different competition mechanisms, the professional division of power construction in developed countries is extremely fine, and the collaborative efficiency of different disciplines has become the most important development content in the construction of substations. Therefore, the efficiency of BIM technology has accelerated the development of power engineering construction units. Introduce [9]. Domestic cable laying software mainly includes GSN developed by Beijing Dameisheng Software Co., Ltd., DLFS developed by Beijing Bochao Times Software Co., Ltd., and Beijing Tianzheng Software Co., Ltd. software. Objectively speaking, there is still a certain gap between these domestic auxiliary cable laying software and PDMS. For example, GSN provides complete digital factory solutions including three-dimensional pipeline design system, equipment design system, and drawing report system [10]. The design software of Beijing Tianzheng Software is usually used in small-scale designed civil buildings; DLFS is used in large-scale power plants, etc. These are in line with domestic engineering specifications and are often restricted in international projects.

This paper mainly studies the substation cable laying schedule based on the GIM three-dimensional model, creates the substation cable laying facility information model of the GIM three-dimensional model, and builds the substation cable laying facility information visualization platform of the GIM three-dimensional model to realize efficient construction management and control.

2 Substation Cable Laying Schedule Based on Gim Three-Dimensional Model

2.1 Research on On-Site Tracking Management Based on 3D Model

Based on GIM 4D technology, the application research of construction schedule optimization and dynamic control is carried out according to the characteristics of power grid projects. In the entire project implementation process, the GIM 4D construction schedule management platform is used to realize real-time data monitoring and information

collaborative sharing, reduce the number of decision-making information transmissions, and ensure that the project is completed on time [11, 12].

- (1) Through mobile terminal equipment and lightweight technology, bring 3D model data and project construction schedule to the construction site, view the plan model in real time through the mobile terminal and record the actual construction data on site to realize the connection between virtual data and actual data.
- (2) Through the three-dimensional model, extract the work list to achieve refined management of materials and resource allocation; carry out accurate schedule simulation of complex construction areas, formulate installation specifications and sequences, and help on-site construction to proceed in an orderly manner.
- (3) Control the construction progress and quality through the collection, integration, calculation and analysis of on-site data; compare the planned status and actual status to understand the progress deviation and potential risk opportunities; analyze the construction efficiency and influencing factors, and carry out the follow-up work Replanning.

2.2 Distance Calculation

The distance between points is divided into two types, Euclidean distance and Manhattan distance.

Euclidean distance: refers to the straight line connection between two points (P1, P2):

$$Dist(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

Manhattan distance: The path between two nodes is at a right angle, the calculation formula of (P1, P2) between the two nodes is as follows:

$$Dist(P_1, P_2) = |x_1 - x_2| + |y_1 - y_2| \quad (2)$$

In the three-dimensional space cable laying, the calculation formula of the path distance between two points:

$$Dist(P_1, P_2) = |x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2| \quad (3)$$

2.3 Requirement Analysis of Early Warning Function of Substation Construction Behavior

At present, relatively backward construction safety management methods and technical means are still used in the construction safety management of substation construction projects. Most of them rely on the formulation of strict system measures and human supervision to increase the probability of casualties or property losses. However, the backward construction safety management methods can no longer meet the new requirements for construction safety in the construction of modern power industry. This paper

combines the characteristics and application characteristics of GIM technology, real-time positioning technology and sensing technology, and adds the required functions to the substation construction safety management system in view of the current status and existing problems of the construction safety management of substation projects. The functional requirements in the system are summarized as follows:

(1) Automated data information collection.

In the construction site of a substation project, there are a lot of data information that needs to be collected, such as the attributes of people, materials, and machines, three-dimensional spatial location information and information on the construction site environment. If only human input is required, the accuracy and timeliness of the data information cannot be guaranteed, nor can it improve the efficiency of safety management at the construction site. Therefore, automated data and information collection is the basis for system implementation.

(2) Real-time updated data information collection.

In the substation construction site, there are a lot of data information that will change with the project construction schedule and site environment changes, such as the three-dimensional position information of people, materials, and machines, the project construction schedule, dangerous areas, and unsafe behaviors. Therefore, real-time tracking and timely updating and sharing of data information can ensure the effectiveness of data information and avoid security accidents in time.

(3) Visual safety monitoring.

Visual safety monitoring can help the relevant personnel of the substation construction safety management to understand and grasp the construction site environment and the working status of the construction personnel in a simple, intuitive and comprehensive manner, and can formulate new safety management plans and goals according to the conditions of the project.

(4) Automated monitoring and early warning.

During the construction of the substation, the construction personnel may be unfamiliar with the construction site environment and construction operation scope, which may lead to safety accidents. Therefore, automated detection and early warning of dangerous areas and unsafe behaviors can make up for the lack of construction safety education and training, prevent construction personnel from accidentally entering dangerous areas, and avoid safety accidents arising therefrom.

(5) Storage and analysis of data information.

The construction of substation projects usually takes a long time. Therefore, the system needs to be able to store data information for a long time, and the positioning data can be stored in the system database to query historical positioning information at any time. The construction data information of the project, including the time period and frequency of unsafe behaviors, are counted as a basis for safety performance and safety education and training, and to improve the safety management efficiency of substation construction projects.

(6) Sharing of data information.

In the substation construction site, the sources of hazards and unsafe behaviors will change with the project construction schedule and the changes in the site environment. Therefore, the sharing of hazard sources and unsafe behavior information

can increase the speed of data information update and improve the efficiency of construction safety management.

3 Compilation of Schedule Plan for Substation Cable Laying Based on Gim Three-Dimensional Model

The project schedule is based on the master node plan made by the owner, and the project master node plan is developed. Then, according to the master node plan, the project plan for the project parts is developed to form the Gantt chart and the network diagram. The content includes the construction project, start time, duration, completion time, labor arrangement, and responsible person of the project site. According to this detailed schedule, weekly inspections and assessments are conducted to form top-down management. At the same time, the grass-roots employees of the project also report the progress plan for the next week in their respective areas in accordance with the actual progress on the spot, forming a bottom-up feedback on the preparation of the progress plan.

The 3D simulation construction function developed based on the GIM three-dimensional model allows the relevant construction management personnel of the project to easily and quickly foresee the progress plan of the project construction. Currently, Bentley's Synchro 4D construction schedule management software is widely used, and Synchro 4D supports a variety of schedule data import, including Microsoft Project and so on. Synchro 4D associates the construction schedule of the Project with the GIM model through the resource allocation in the software (the corresponding model component can be selected for each sub-item construction), and can also export the schedule plan of the resource task set in Synchro to the Project file for adjustment. After the construction is completed, it is synchronized to Synchro 4D, the related resource tasks are automatically changed, and the 4D simulation construction animation is reflected in the 4D review in Synchro 4D.

In view of the fact that the search path of the ant colony in nature is tree-shaped, which is consistent with the layout of the cable in the cable laying; at the same time, the ant colony algorithm is a discrete optimization algorithm, which is a probability algorithm suitable for finding the optimized path in the graph. The ant colony algorithm is selected and the ant movement rules are modified accordingly. In this paper, simulation examples and a large number of parameter studies have verified the validity of the data.

4 Experimental Analysis of Substation Cable Laying Schedule Research Based on GIM Three-Dimensional Model

4.1 Substation Layout Design Based on GIM Three-Dimensional Model

In the design code, the design rules for substation layout can be summarized into three aspects:

- 1) Working environment: electrical equipment should be arranged in a reasonable place and the surrounding natural environment should meet the requirements for safe operation of the device.

- 2) Fire protection: fire protection requirements should be considered in the layout of electrical equipment, and necessary fire protection facilities should be provided.
- 3) Maintainability: Sufficient space should be reserved between electrical equipment, and between electrical equipment and buildings (structures) to meet maintenance requirements.

The review, reasoning and calculation of the above three types of layout design rules mainly require the support of three types of calculation functions: basic algebraic operations, spatial analysis, and geometric model calculations.

Taking the “3 ~ 110 kV High Voltage Power Distribution Device Design Specification” as an example, the basic calculation function requirements involved in various layout rules are shown in Table 1.

Table 1. Basic calculation function demand statistics

Design rule type	Basic calculation function		
	Algebraic operations	Space Analysis	Geometric calculation
Working environment	8	4	3
Fire protection	6	18	19
Maintainability	8	12	14

Generally speaking, spatial analysis and geometric model calculations are the focus of rule coding. They all involve more complicated geometric (shape) calculations or spatial analysis algorithms. Such algorithms are not easy to describe in logical language; although hard coding can also be used to translate, they require high programming skills of design experts and difficult coding.

The programming framework after visualizing the layout rules of substation construction is shown in Fig. 1.

The bottom layer is the basic layer, which contains predefined various nodes, terminals, and streams. The node not only encapsulates the aforementioned three basic calculation functions, but also encapsulates functions such as design model access. The terminal is the data input and output interface of the node, and the flow is the way to realize data transmission and flow control, including two types of data flow and control flow. The cable laying system is mainly the connection between two equipment nodes, but under special circumstances, the connection between multiple equipment nodes will be carried out. In this paper, considering the diversity of cable connection methods and the tree connection structure between device nodes, with the help of the algorithm's advantage in finding the optimal tree path, an algorithm-based cable laying system simulation software is developed, which can support multiple device nodes. The optimal path between the two can be searched for the optimal path better.

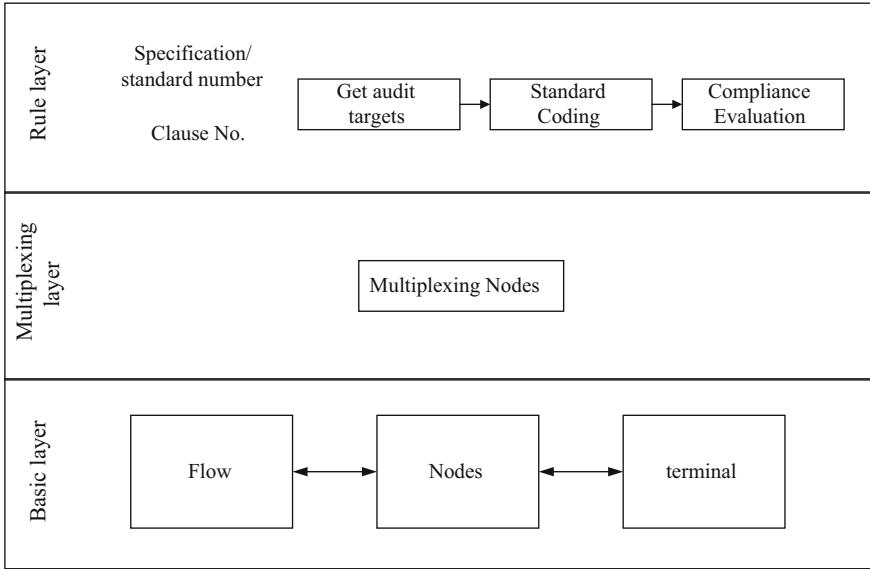


Fig. 1. Visual programming framework for design specifications

4.2 Progress Monitoring

In view of the similarity in cable laying between substations and power plants, the cable laying of substations can be used as an example. Take the Yuxiang substation as an example. The substation has three power distribution devices with voltage levels of 110 V, 220 V, and 330 V, all of which are connected to the equipment in the main control room. This paper uses an ant colony optimization algorithm to simulate the optimization of the cable laying path from the main control building of the substation to the power distribution equipment on its side. The GIM three-dimensional model contains building information such as component size, quantity and material, and time is added on this basis. The information forms a 3D model. The entire construction process can further analyze and understand the design drawings and construction plans, and evaluate the possible conditions during the construction and the factors affecting the progress, so as to formulate countermeasures and optimize the construction plan in depth. The process is shown in Fig. 2.

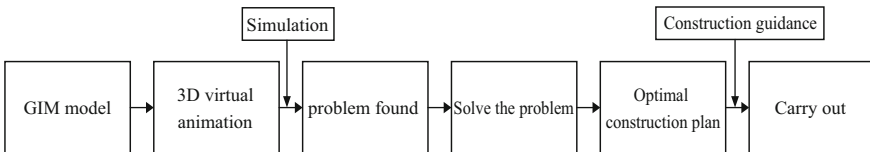


Fig. 2. Flow chart of progress management

GIM-based schedule management is based on traditional schedule management theories and technical methods, combined with GIM related technologies and the project schedule information management platform, to realize the construction simulation of the project schedule. The main way of GIM-based schedule management is to build a GIM-based schedule management platform. With the GIM information platform as the core, the relationship between GIM model, WBS and schedule information is established, and finally real-time monitoring and dynamic simulation of project schedules, as well as engineering quantities Dynamic control, etc. This method is conducive to the adjustment and optimization of the schedule, and greatly improves the schedule management process.

5 Conclusions

With the continuous development of computer software and hardware technology, China's transmission and transformation engineering design is undergoing a development transition from a two-dimensional era to a three-dimensional era. GIM three-dimensional design is a major innovation in design methods. GIM three-dimensional design is becoming an inevitable development trend in the design of power transmission and transformation engineering industry, and relevant units have begun to explore gradually. Substation is one of the important links of power supply, and its construction is a huge and complex project that requires the coordination and cooperation of different professionals. The application of enterprise GIM technology should be more in-depth, not just on some primary applications such as collision detection. The value of GIM technology lies in its information. Only with the active participation of all parties can GIM technology truly blossom in China.

References

1. Abidin, A., Mustafa, S., Aziz, Z.A., et al.: Subsea cable laying problem. *Matematika* **34**(2), 173–186 (2018)
2. Benato, R., Sessa, S.D., Forzan, M., Marelli, M., Pietribiasi, D.: Core laying pitch-long 3D finite element model of an AC three-core armoured submarine cable with a length of 3 metres. *Electr. Power Syst. Res.* **150**, 137–143 (2017)
3. Rui, L., Fan, X., Yin, X., et al.: Cable target image segmentation method for manual cable laying process. *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/J. Comput.-Aided Design Comput. Graph.* **30**(4), 666 (2018)
4. Guan, L., Wang, W., Junhui, H., Li, Z., Wang, H., Yang, X.: Fineness standard and implementation method of 3D model of electrical equipment in substation (Converter Station). *IOP Conference Series: Earth and Environmental Science* **428**(1), 012074 (2020)
5. Wang, Q.-X., Zhou, Q.-P., Zhou, L., Tang, X.-J., Ke, F.-C., Sang, Z.-X.: The theoretical research on modelling method for the grid information model of substation. *IOP Conference Series: Earth and Environmental Science* **587**(1), 012085 (2020)
6. Ren, C., Wang, Y., Xue, K., Zhi, S., Chen, T.: Research on application of construction model based on 4D visualization in construction of intelligent substation. *IOP Conference Series: Earth and Environmental Science* **632**(4), 042028 (2021)

7. Qiaoyun, W., Yang, H., Wei, M., Remil, O., Wang, B., Wang, J.: Automatic 3D reconstruction of electrical substation scene from LiDAR point cloud. *ISPRS J. Photogramm Remote Sensing* **143**, 57–71 (2018)
8. Guan, L., Wang, W., Junhui, H., Zhang, Y., Wang, H., Yang, X.: Research on general model library framework of substation engineering. *IOP Conference Series: Earth and Environmental Science* **428**(1), 012087 (2020)
9. Junhui, H., Li, Z., Jing, T., Hong, M., Mingxun, H., Wenjie, X.: Research on lightning protection verification based on three-dimensional model of substation. *IOP Conference Series: Earth and Environmental Science* **770**(1), 012023 (2021)
10. Gao, Y.: Application of 3D design technology in substation design. *IOP Conference Series: Materials Science and Engineering* **782**(3), 032086 (2020)
11. Wang, Y., Chen, M., Wang, T., et al.: Research on 3D visual component assembly technology of smart substation simulation. *Dianli Xitong Baohu yu Kongzhi/Power Syst. Protection Control* **46**(023), 118–126 (2018)
12. Guo, Y., Yuan, H., Zhuang, Y., Jie, X., Zhang, Y.: Research on informatization construction of digital substation in smart grid. *IOP Conference Series: Earth and Environmental Science* **632**(4), 042089 (2021)



Social Service Management System Based on Logistic Regression Algorithm

Huajun Li^(✉)

Jiangnan University, Wuhan 430056, China

Abstract. The use of a sound social service management system to improve the level of grassroots social governance has become an inevitable trend to strengthen grassroots social service management. The purpose of this article is to study the social service management system based on the Logistic regression algorithm. First, it introduces the generalized linear model of logistic regression, and then describes the terminal functions and management requirements of the social service management system. Designed and implemented a social service management system based on Logistic regression analysis. Through the visualization of data, the department can understand various information more intuitively when applying the system, and provide an intuitive management method. Logistic regression algorithm was used for the first time in the field of urban management and poverty reduction, and gradually extended to the basic public service department. Practice has proved that the social service management based on the logistic regression algorithm is effective in maintaining social harmony and stability, promoting the construction of service-oriented government, and mobilizing society. Significant results have been achieved in terms of force participation in social governance.

Keywords: Logistic regression algorithm · Social service · Management system · Social governance

1 Introduction

The scale of data continues to expand, and machine learning tasks face more challenges [1]. Therefore, simple logistic regression has received more attention [2]. How to accelerate economic development, promote social stability, and standardize social behavior has become a common problem faced by people across the country [3]. Similarly, how to use information technology to manage the society's services has become the first problem to be solved [4]. In order to meet the trend of merger and promotion of domestic social development, it is necessary for Weifang City to use advanced network technology to design the social service management system, realize the society as a carrier, rationally plan resources, and coordinate social development and promote steady economic growth. Reduce development risks, etc. to provide a good foundation [5].

Social service is not only one of the main functions of higher vocational colleges, but also the most direct manifestation of the core competitiveness of higher vocational colleges. According to the characteristics of education and teaching of management and

service majors, some scholars expounded the construction of social service system in management and service colleges and universities, and put forward some suggestions and countermeasures for the construction of social service system in such colleges and universities. Colleges and universities will play a certain role in promoting the development of higher vocational education [6]. Mustafa A proposed a model based on the combination of disordered multinomial Logistic regression (MLR) and cellular automata (CA) to simulate the construction of expansion and densification. The possibility of completion and development is assessed based on (i) a set of causal factors for completion and development and (ii) the land use of adjacent units. The model considers four built-up areas: non-built-up areas, low-density, medium-density, and high-density built-up areas. Unlike the most commonly used building/urban models that simulate the expansion of buildings, their method takes into account the potential for expansion and densification in the built area where the existing density allows. The model was constructed, calibrated and verified for the Wallonia region (Belgium) using cadastral data [7].

At present, there is still a blank in the domestic academic circles regarding the social service management system based on the Logistic regression algorithm. According to the search on CNKI, most scholars' research content is mainly about the dilemma of social service management and the situation of supply and demand, the forecast of the development trend of old-age care, and the exploration of old-age care in social security. From the perspective of Logistic regression algorithm, this paper deeply explores the needs of social service management, analyzes the basic situation and social status of M area, and predicts the probability of returning to poverty, constructing a reasonable and feasible social service system that meets the needs of M area. Open up a new direction for the research of social service in the future.

2 Research on Social Service Management System Based on Logistic Regression Algorithm

2.1 Logistic Regression

Logistic regression is a generalized linear model [8]. The regression problem is a curve adjustment process. In fact, the curve found is not randomly selected [9]. Generally, the shape of the curve is assumed first, such as a straight line, a square, etc., and then various parameters are obtained through calculation and learning [10].

There are two problems in the parameter learning process: using straight lines to customize nonlinear problems will cause major errors [11, 12]. The other is that the required parameters are not ideal, and there will be insufficient or over-adjustment. Generally, the selected model conforms to the characteristics of the data, especially in the processing of high-dimensional data, the data is generally linearly separated.

The role of the loss function is to evaluate the quality of the model. Commonly used loss functions include 0–1 loss function, quadratic function and so on.

2.2 Terminal Functions and Management Requirements of the Social Service Management System

When analyzing the functions of mobile terminals, analyze the functional requirements of the terminal from the aspects of data information, interaction and services. Through

the terminal system, the location, population, and regional information can be managed, and important information such as meetings and documents can also be managed. In addition, the log recording function is also required in the terminal function to perform the functions of the relevant personnel record.

In addition, end users can also use the interactive communication service function to publish information in the form of short messages, read relevant information on the platform, and understand relevant information in a timely manner.

The functions in the terminal function management business include basic information management, work service management, work log function and interactive communication function. The detailed introduction of each function is as follows.

- (1) Basic information management The basic information management function includes the management of grid area information, people's daily information, and social organization information. In this function, the maintenance and query functions of population-related information in the jurisdiction are provided.
- (2) Work log function Relevant staff can use this function to record important work situations. At the same time, the work log will also be used as an assessment standard for employees. Various services such as work log selection can be performed in this function.
- (3) Service service management The service service management function is related to events acceptance, filing and handling. Important meetings, documents, work reports, etc. can be managed in this function.
- (4) Interactive communication function Through the interactive communication function, business personnel can obtain information through platforms, text messages, etc., and can also view important meetings, documents and other materials.

3 Investigation and Research on Social Service Management System Based on Logistic Regression Algorithm

3.1 Experimental Environment

In the process of choosing a programming language, JSP is the preferred programming language for the development of the social service management system platform of this article. It has more frameworks and components, which improves the development efficiency of the system platform. The cross-platform feature of JSP can make it run in various types. In the server, the stability of the system is improved, which makes the social service management system platform of this article easier to maintain. The well-known relational database Oracle is selected to be responsible for information interaction in the social service management system platform.

3.2 Logistic Regression Model

The general classification rule of the logistic regression algorithm is based on the output value of the sigmoid function as the threshold, and 0.5 as the threshold. Its corresponding decision function is expressed as:

$$y = 1, \text{if } p(y = 1|x) > 0.7 \quad (1)$$

For the selection of the threshold, sometimes the required threshold is selected according to the actual situation. For example, when the judgment of positive examples needs to be reliable, the threshold can be selected slightly larger.

The basic idea of the stochastic gradient ascent algorithm is: look along the gradient direction of this function, and the result found is the maximum value of the function we need to solve. If the gradient is denoted as ∇ , then the gradient representation of the function $f(x,y)$ is:

$$\nabla f(x,y) = \begin{pmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{pmatrix} \quad (2)$$

This means that the variable values x , y move in their respective directions.

The functional modules based on Logistic regression analysis mainly include: introduction to poverty return prediction model, data display module of poverty alleviation households, and data display module for predicting return to poverty.

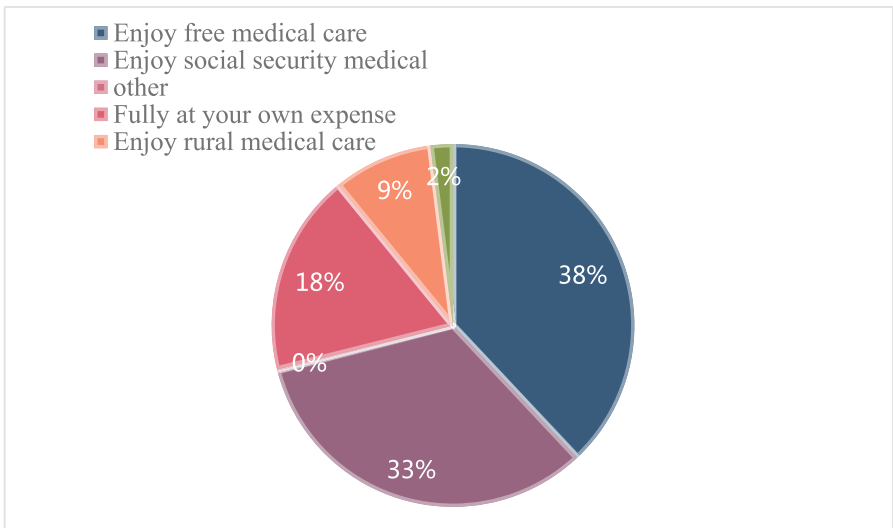
4 Investigation and Research Analysis of Social Service Management System Based on Logistic Regression Algorithm

4.1 Basic Information Module

The information chart is calculated according to the data flow of the outline design and detailed design stage. Enter the detailed information of the M area in the column. The basic situation overview of the M area is intuitively displayed on the page. Here is a virtual table of data generated according to the household registration category, namely It exists in the view in Oracle, which guarantees the security of the data. Select the corresponding street and community to query the information that needs statistics, and click the pie-shaped legend of the query result to enter the detailed information of the household registration status. In the cascading menu for selecting streets and communities, JavaScript technology is used. The household registration category is the ID value saved in the page and is directly transferred to the household registration category detail page by clicking. Select the corresponding street and community and the information that needs statistics. Perform a query and click on the pie legend of the query result. You can enter the detailed medical insurance information as shown in Table 1. You can clearly see the relevant medical insurance situation in the pie chart, as shown in Fig. 1, in the three-level linkage menu. With the cooperation, the data is returned to the page by the select query statement:

Table 1. Medical statistics

	Percentage
Enjoy free medical care	38
Enjoy social security medical	33
Other	0.1
Fully at your own expense	18
Enjoy rural medical care	8.9
Enjoy half-fee reimbursement from relatives	2

**Fig. 1.** Medical statistics

The above information is queried with the household registration database as the basic database. In the outline design and detailed design, the user registration information between the various modules of the system platform is generally used as the basic data, and then each department is constructed according to the situation of each department. Data, thereby realizing the design principle of “high cohesion, low coupling” of the system, which not only improves the efficiency of development, but also makes the maintenance of the system platform easier.

4.2 Poverty Return Prediction Module

The display module of the return-poverty prediction model is mainly introduced on the home page of the model, explaining the basic functions implemented by the module and the related algorithms and processing procedures adopted by the poverty alleviation

department, including the functions implemented by the model, data preprocessing, and logistic regression algorithms. Introduction, cluster analysis, etc., so that the staff of the Poverty Alleviation Office and non-professionals can understand the function of the model when applying the model. The poverty alleviation data page is mainly for the visualization of the poverty alleviation data of Gansu Province, so that the poverty alleviation department can have an intuitive understanding of the current poverty alleviation situation in M area through this module. Among them, the newly added poverty-stricken households are made by supporting poor households to become poverty-stricken households on the basis of previous years. The number of people who have been lifted out of poverty is the number of people who have been lifted out of poverty before that year.

Among them, the module ability index compares the average value of various indicators in the poverty alleviation households with the overall average value of the poverty alleviation households. The difference is the ability index. The main reason for poverty is that the household counts the reasons for poverty in the poverty alleviation survey. By clicking on the household data, you can intuitively see the basic information of the household.

The Poverty Forecast Data page mainly displays the Poverty Poverty Forecast data in Region M, as shown in Fig. 2.



Fig. 2. Display of M-regional information data for the prediction of returning to poverty

Among them, the meaning of green stable poverty alleviation is whether there is a phenomenon of returning to poverty before the households that have not returned to poverty in the forecast. If there is a sign of returning to poverty under the household characteristics, if the household has signs of returning to poverty before, after subsequent support, the poverty-stricken households that are predicted to return to poverty are called poor households, and there is no sign of returning to poverty before the orange sign of poverty. For returning to the poor.

The core page of the poverty return prediction model includes the prediction results of the population returning to poverty in 2022 in the M region. There are six clusters. The probability of returning to poverty is predicted by logistic regression, and the reason analysis shows the key indicators under the corresponding category of the predicted returning to poverty.

5 Conclusions

At present, innovative social service management has been included in the important tasks and agendas of party committees and governments at all levels. The social service management system is still at the stage of strengthening the government's social management and social control capabilities. In addition, urban and rural people have higher expectations for improving the level of social services and satisfying diversified service needs. Therefore, it is necessary for the government to re-examine the social service management system from the perspective of governance, making it an important carrier and platform to promote the innovation of grassroots social service management, and then improve the level of grassroots social service management. Therefore, this thesis is supported by the social service management system, based on the domestic and foreign theories and local practices, taking the specific practice of M area as an example, to create a social service management system based on the Logistic regression algorithm. Analyze the results and existing problems of social service management system construction, and explore methods and measures to promote the in-depth development of social service management system based on Logistic regression algorithm from the perspective of innovative grassroots social service management.

References

1. Liang, L., Yunsheng, P., et al.: An accelerator for the logistic regression algorithm based on sampling on-demand. *Science China (Information Sciences)* **63**(06), 226–228 (2020)
2. Yu, Y., Xiong, Z., Xiong, Y., et al.: Improved logistic regression algorithm based on kernel density estimation for multi-classification with non-equilibrium samples. *Comput. Materials and Continua* **61**(1), 103–118 (2019)
3. Lei, D., Tang, J., Li, Z., et al.: Using low-rank approximations to speed up kernel logistic regression algorithm. *IEEE Access* **PP**(99), 1–1 (2019)
4. Feng, L., Jindong, W., Huang, J., Qiu, X.: Aircraft engine degradation prognostics based on logistic regression and novel OS-ELM algorithm. *Aerospace Sci. Technol.* **84**, 661–671 (2019)
5. Antonio, P.B., Manuel, F., Ernesto, C.C., et al.: Sample size calculation to externally validate scoring systems based on logistic regression models. *Plos One* **12**(5), e0176726 (2017)
6. Ksiazek, W., Gandor, M., Plawiak, P.: Comparison of various approaches to combine logistic regression with genetic algorithms in survival prediction of hepatocellular carcinoma. *Comput. Biol. Medicine* **134**, 104431 (2021)
7. Mustafa, A., Heppenstall, A., Omrani, H., Saadi, I., Cools, M., Teller, J.: Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. *Comput. Environ Urban Syst.* **67**, 147–156 (2018)
8. Nishimoto, T., Mukaigawa, K., Tominaga, S., et al.: Serious injury prediction algorithm based on large-scale data and under-triage control. *Accid. Anal. Prev.* **98**, 266–276 (2017)

9. Divya, R., Shantha Selva Kumari, R.: Genetic algorithm with logistic regression feature selection for Alzheimer's disease classification. *Neural Comput. Appl.* **33**(14), 8435–8444 (2021)
10. Robles-Velasco, A., Cortés, P., Muñuzuri, J., Onieva, L.: Estimation of a logistic regression model by a genetic algorithm to predict pipe failures in sewer networks. *OR Spectrum* **43**(3), 759–776 (2021)
11. Hrynzovskyi, A.M., Holovanova, I.A., Omelchuk, S.T., et al.: Public health and social supervision issues within public administration of Ukrainian territories in the late 8th- early 9th centuries. In: *Wiadomości lekarskie*, vol. 71, (1 pt 2), pp. 246–251. Warsaw, Poland: 1960 (2018)
12. Mihi, S., Ben Ali, B.A., El Bazi, I., Arezki, S., Laachfoubi, N.: Multi-way Arabic Sentiment Classification Using Genetic Algorithm and Logistic Regression. In: Hassanien, A.E., et al. (eds.) *AICV 2021. AISC*, vol. 1377, pp. 317–325. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_29



Timing System of Track and Field Competition Based on Data Analysis Algorithm

Yingyue Hu^{1(✉)} and Sinem Alturjman²

¹ Physical Education School, Wuhan Business University, Wuhan, China

² Near East University, Nicosia, Turkey

Abstract. Today, in the context of the big data era, increasingly advanced data processing technologies have been widely used in various fields for data analysis. At the same time, people are paying more and more attention to the standardization of competitions and the fairness of results in sports events. Data analysis is reflected in sports competitions. For example, coaches analyze training or competition data to help athletes improve their competitive level, formulate competition strategies, and optimize the configuration of team athletes. Through the analysis, exploration, display and prediction of sports data, it provides value to data users. At the same time, the track and field timing system automatically records and automatically classifies, archives and displays the performance of each athlete, making the timing method more efficient and convenient, and also plays an important role in making the game more accurate and fair and reducing labor costs. The automatic timing system for track and field competitions monitors the athlete's time spent in the competition and the distance from the finish line in real time through wireless communication between the base station at the finish line and the intelligent terminal. At the same time, combined with data analysis algorithms, a more convenient and safer and more reliable automatic timing method is used to determine the conditions for athletes to reach the end point, so as to achieve the goals of accurate ranging, cost reduction, and error reduction. This article adopts experimental analysis and data analysis methods to explore the timing system that is more suitable for track and field competitions by understanding the algorithms related to data analysis, so as to achieve accurate ranging, reduce costs, and reduce errors. According to the research results, the response time of the system is kept within 5 s. The performance of the system tested this time is still very good, and it has certain application value.

Keywords: Data analysis · Track and field competition · Game timing system · Image recognition

1 Introduction

Regardless of type or quantity, the era of information explosion has brought unprecedentedly huge information data to people. However, how to effectively select data and how to obtain useful information from data analysis has become an important issue for

researchers. At the same time, in sports events, the regularity of the game and the fairness of the results have attracted more and more attention and attention. Therefore, it is very important to emphasize the fairness of the game and reduce the impact of manual intervention on the results of the game. In track and field competitions, the accurate judgment of athletes' performance is more dependent on high-precision timing results. Therefore, it is of great significance to combine data analysis technology to develop a more practical, accurate and efficient track and field game timing system. While it can more accurately measure the results of the game, it also provides convenient, fast and stable services for the development of mass sports.

From the current point of view, the research results on data analysis and processing algorithms are relatively rich, while the research on the timing system of track and field competitions is relatively small. For example, Song Bodong proposed that distributed big data analysis and processing algorithms can effectively improve the speed of the system in analyzing large amounts of data, and have high application value in high-frequency big data processing systems [1]. Chen Yuanyuan believes that data processing algorithms can not only compress data, but also help extend the life cycle of the network and reduce network energy consumption [2]. Wang Tao pointed out that the new race timing system and the application of automation technology in the new system effectively make up for the shortcomings of the current system. At the same time, it also showed people the future development direction of the new track timing system [3]. Therefore, this article is based on the data analysis algorithm to study the timing system of track and field competitions, which has certain research significance and application value.

This article mainly discusses these aspects. First, a brief description of data analysis and related knowledge is given. Then it briefly elaborated on the timing system of track and field competitions. The application of image edge recognition algorithm in track and field competition timing system is also introduced. Finally, based on the data analysis algorithm, an experimental research on the timing system of track and field competitions is carried out, and the experimental results and analysis conclusions are drawn.

2 Analysis of Track and Field Competition Timing System Based on Data Analysis Algorithm

2.1 Data Analysis

The level of data quality often determines the quality level of data analysis and data mining. Because the messy data cannot be efficiently analyzed and mined, the data preprocessing technology has emerged at the historic moment. Preprocessing the data before data mining can improve the integrity, accuracy and usability of the data, remove redundant data, reduce complex calculation operations, and improve the quality of data mining.

More effective analysis of data has become an inevitable demand in many fields. Having hidden value is one of the characteristics of data, and data analysis algorithms are one of the important ways to mine its hidden value. Among data analysis algorithms, machine learning and data mining algorithms are a large class of core algorithms, used for various types of big data analysis applications, and therefore have received extensive and continuous attention and research from the industry and academia.

Only after the massive data is analyzed and processed, can the useful information contained in the massive data be obtained. At present, more and more socio-economic and technological research fields involve the application of big data, so the method of big data analysis becomes particularly important. It can be said that the key factor in determining whether the information is valuable is whether the analysis method is used properly. Generally speaking, data analysis is carried out from these two aspects. One is visual analysis. Visual analysis can show their attributes through intuitive data graphics, and users can better understand the complex relationships behind the data through comparable information. The second is data mining algorithms. This algorithm generally conducts in-depth data mining based on the type and format of the data, and extracts the identified useful information. Because the amount of big data is very large, these data mining algorithms provide faster solutions for big data processing [4, 5].

Combined with data analysis algorithms, the key events of athletes in training or competition are recorded as structured data, and then data processing technology is used for statistical analysis of the data. The scope and depth of these data is very limited. With the development of data acquisition technologies such as cameras and portable devices, the data collected has become more extensive and continuous. These data can fully capture complete information about training or competition. However, this also brings difficulties and challenges to data analysis and mining. The collected continuous raw data must be converted into structured data before the data can be effectively analyzed [6, 7].

2.2 Timing System for Track and Field Competitions

Track and field competitions are an important part of major sports events, including numerous events, scattered venues, and factors such as the number of participating teams, athletes, officials, the number of matches, the number of referees, and the large amount of information transmitted and scattered information. In general, large-scale track and field events include three events, while the number of applicants for sprint events can reach four events. Even the field events are generally divided into two events, qualifying events and final events. Therefore, the competition management of large-scale track and field events must be accurate, efficient, standardized, and simple, and facilitate the objective extraction of the results of the on-site processing of the competition. It is often operated by humans alone. It is difficult for competition management to achieve rapid, accurate and timely information disclosure requirements. There is no doubt that it is very necessary to design a complete competition timing system [8, 9]. The composition of the entire timing system is shown in Fig. 1.

From the current point of view, in track and field competitions, timing methods mainly include traditional manual timing, optical timing and radio frequency identification timing. However, due to the large influence of human operation and low accuracy, traditional manual timing is mainly used for informal track and field competitions. Compared with manual timing, the optical timing method improves the accuracy to a certain extent, but it will still be affected by manual intervention.

Generally speaking, the first thing that appears is a fully automatic electronic timing system, which consists of a high-speed camera system and an electronic timing instrument. The film synchronously records the image and time data of the athletes, and then

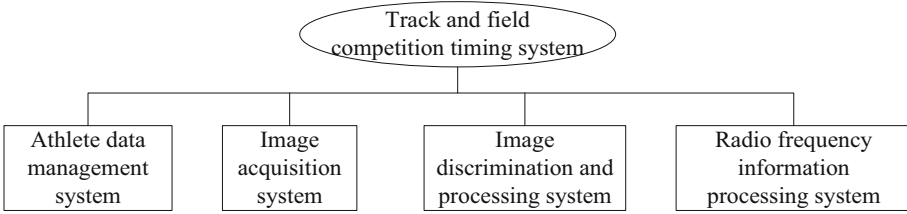


Fig. 1. Diagram of Timing System Composition

the referee determines the athlete’s ranking and performance based on the recording results on the film. However, the system needs to transport a large amount of film, film production and interpretation speed is long, image clarity is also difficult to guarantee, and film and other items are expensive, and the cost is high, which increases the cost of timing track and field competitions. The second is the photoelectric CDD sensor, which collects the narrow slit image at the end point, and after image processing, time scale superimposition, and image reconstruction, a continuous narrow slit image with time stamps is formed and stored on the video tape. These images are transmitted from the video recorder to the computer, and the computer is used to digitally process the images, and then the referee evaluates the position and result of the athletes at the finish line by moving the vertical cursor. Finally, the most advanced electric timing system at present has a higher accuracy of image interpretation, which can effectively distinguish the position of each athlete to the finish line and accurately determine the athlete’s competition performance [10, 11].

2.3 Application of Image Edge Recognition Algorithm in Timing System of Track and Field Competition

The image edge recognition algorithm detects the information contained in the image by the edge of the image through recognition, classification and processing. It is one of the basic methods for advanced image technology such as image analysis and processing, object feature extraction, and machine vision. As the core of image processing, compared with using the processing software on the main control computer, the image processing time has been reduced many times. At the same time, the main control computer has also been relieved, and the interpretation speed has been accelerated. The most effective way to improve the accuracy of interpretation is to improve the accuracy of the recognition of the edge of the image. Since the target edge is a fuzzy concept in the image, that is, the part that changes drastically, the accurate detection of the target edge is directly related to the accuracy of the measurement. The edge recognition algorithm requires first overall interpolation, then determine the target image curve, and finally calculate the edge points to accurately identify the edge of the image, so as to use the image for score interpretation, improve the interpretation accuracy, and achieve the high accuracy of the track and field game timing system [12]. The calculation method is as follows.

$$T_u(y) = b_{u3}y^3 + b_{u2}y^2 + b_{u1}y + b_{u0}y \in [y_u, y_{u+1}] \tag{1}$$

$$T_u(u + 1) = G(u + 1) \tag{2}$$

Among them, u is the pixel, y is the edge point, $Tu(y)$ is the piecewise cubic function, and $G(u)$ is the gray value of the u -th pixel, and $\{u, G(u)\}$ is the pixel point.

3 Experimental Research on Track and Field Competition Timing System Based on Data Analysis Algorithm

3.1 Experimental Background

Testing system performance is an indispensable part of software testing, and it is also a process of verifying whether system performance indicators meet user needs. The system performance test is usually performed after the functional test is completed. In order to ensure the integrity of the function, the system realization process has been optimized so that the system can achieve the required performance in demand modeling.

3.2 Experimental Environment

In this experiment, the track and field game timing system will run on the wide area network and use the Internet based on the TCP/IP network protocol. The current operating system of the database server will adopt Windows Server 2018. The database uses SQLite for management operations. The client hardware environment adopts a PC with dual-core CPU, Inter Core i5 processor, 3GHz main frequency, 800GB or more hard disk, 8GB memory, and independent graphics card. The network configuration uses 1000Mbps wide as wired.

3.3 Experimental Process

In this experiment, the system performance was tested. The test content mainly included the system's concurrency, carrying capacity, response time, server utilization, and CPU occupancy. This experiment achieves the goal of real-time performance monitoring by simulating the concurrent operation behavior of actual users. In this test, 50, 100, 150, and 300 people were simulated to enter and use the system at the same time, and then the required test results were obtained. Finally, the data was analyzed to understand the performance of the system. Some test results are shown below.

4 Experimental Analysis of Track and Field Competition Timing System Based on Data Analysis Algorithm

This experiment was carried out under the condition that the network environment is normal and all the data are at standard values. It simulates 50, 100, 150, 300 people entering and using the system at the same time, and the response time, server utilization, and CPU usage of the system Carry out the test, get the corresponding data result, so as to understand the performance of this system. The test results are shown in Table 1.

It can be seen from Fig. 2 that when 50, 100, 150, and 300 people are simultaneously accessing the track and field game timing system in this experiment, the system reaction

Table 1. Test Results of System Performance

Number of concurrent users	Response time (seconds)	Server utilization rate (%)	CPU usage(%)
50	1.46	1.7	3.1
100	1.98	2.3	3.5
150	2.56	2.8	3.8
300	3.29	3.4	4.3

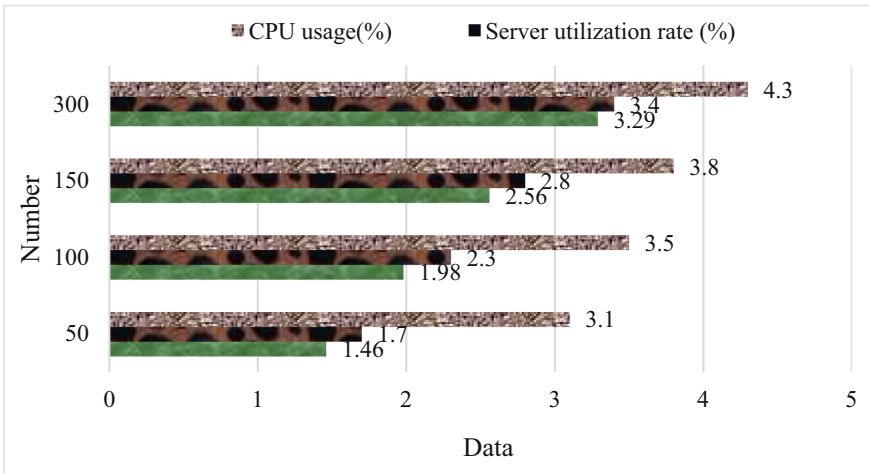


Fig. 2. Test Results of System Performance

time is 1.46, 1.98, 2.56, and 3.29s respectively, and the CPU occupancy rate has also increased from 3.1% to 4.3%. It can be seen that the system response time is kept within 5s, and the performance of the system tested this time is still very good and has certain application value.

5 Conclusion

In the current era, with the rapid development of information technology and the expansion of various data scales, more and more advanced data processing technologies are used in various fields of data analysis. At the same time, in order to respond to the call for scientific and standardized development of sports games, it is a very critical step to explore the game timing system. Because the track and field game timing system can replace manual timing and minimize manual intervention, it is helpful to determine the fairness and justice of the game, and it also improves the level of athletic events, and accelerates the scientific and standardized process of sports games. Therefore, this

article uses the advantages of data analysis algorithms to analyze and research track and field competition timing systems and related content, which meets the requirements of fairness and scientificity in track and field competitions, and has important practical significance.

References

1. Song, B., Zhang, L., Jiang, Q.: Research on distributed big data analysis algorithm based on Spark. *Comput. Applications and Software* **36**(01), 45–50 (2019)
2. Chen, Y., Zhu, X.: Sensor network data processing algorithm based on model fitting. *J. Zhoukou Normal Univ.* (2015-2), 66–70 (2021)
3. Wang, T.: The design of a new track timing system. *Bull. Sports Sci. Technol. Lit.* (2014-2), 29–30 (2021)
4. Rebeiro, C., Mukhopadhyay, D.: A formal analysis of prefetching in profiled cache-timing attacks on block ciphers. *J. Cryptol.* **34**(3), 1–39 (2021)
5. Spolon, A., Zampieri, L., Burtovoi, A., et al.: Timing analysis and pulse profile of the Vela pulsar in the optical band from Iquye observations. *Mon. Not. R. Astron. Soc.* **482**(1), 175–183 (2019)
6. Leede, E., Buchanan, F.R., Coopwood, T.B., et al.: Is faster better? An aortic trauma foundation database analysis of aortic injury repair timing. *J. Am. Coll. Surg.* **231**(4), S320–S321 (2020)
7. Aleshkin, A.P., Zybarev, K.K., Ivanov, D.V., Makarov, A.A.: A method for positioning and timing support of remote users based on mobile radio interferometry. *Gyroscopy Navig.* **11**(1), 77–84 (2020)
8. Karamanidis, K., Epro, G.: Monitoring muscle-tendon adaptation over several years of athletic training and competition in elite track and field jumpers. *Front. Physiol.* **11**, 607544 (2020)
9. Campbell, E., Poudevigne, M., Irving, R., et al.: Psychophysiological stress and performance in Jamaican junior track and field athletes. *Perform. Enhanc. Health* **8**(2–3), 100171 (2020)
10. Brajkovic, V.: From one track-minded to multimodal. *Progress. Railr.* **62**(6), 31–34 (2019)
11. Garcia, F., Vázquez-Guerrero, J., Castellano, J., et al.: Physical demands between game quarters and playing positions on professional basketball players during official competition. *J. Sports Sci. Med.* **19**, 256–263 (2020)
12. Nandu, J., Noordin, H.: An evaluation on the pre-competition anxiety levels of under eighteen male track and field athletes from the rural areas of Sabah. *Malaysian J. Movement Health and Exercise* **9**(1), 141–148 (2020)



Financial Software Perception Function Based on Big Data Intelligent Technology

Hongyan Li^(✉)

Dianchi College of Yunnan University, Kunming 650000, Yunnan, China

Abstract. Financial management is an important part of modern corporate governance, and the study of financial management information has also become an important subject of financial software research. In order to adapt to the frequent changes of the open environment and the needs of users, The financial system must be able to continuously adapt to changes in the internal operating environment of the software system and changes in external functional requirements. The purpose of this article is based on intelligent information technology and the rapid development of the Internet, apply intelligent information technology to the development of financial software perception functions, and designing the perception function of financial software to ensure the functional aspects of new financial software progress. This paper takes the perception function of financial software based on intelligent confidence technology as the research goal. With the support of big data intelligence data mining technology, the data model and algorithm studied in this paper are given, namely the Logistic algorithm and ID3 algorithm in data mining technology. Then the two algorithms are used to verify the data trends of a certain securities website on the Internet for two years as variables. Data experiments are used to prove the feasibility and practicability of the algorithm proposed in this paper. According to the data experiments in this article, The feasibility of the financial software perception function of the intelligent information technology proposed in this paper, substitute the real data of a website for two years into the data mining algorithm to get, which is based on the Logistic algorithm and the ID3 algorithm of the decision tree. The prediction accuracy of both exceeds 80%, and the accuracy of the decision tree algorithm reaches about 86%, which is better than the Logistic algorithm.

Keywords: Big data technology · Intelligent technology · Financial software implementation · Perception function

1 Introduction

In recent years, with the rapid development of the Internet of Things and big data technology, big data technology has become a key technology of global concern. The ability to analyze and use large amounts of data has become a new standard for measuring the resilience of a country. Intelligent technology in big data has become the source of competition between countries and enterprises. All these phenomena indicate that the “big data era” has arrived. In general, in the last few years of the big data era,

the transformation of information technology has gone through various stages such as “decision support, data management, data warehouse, data mining and big data era”.

In the research on the perception function of financial software based on big data intelligence technology, Some experts have explored this direction and obtained good research. The application of intelligent information technology and data processing technology in all aspects of life has progressed While leading in information technology, it has also generated many information security problems. Explore the information security system and data processing technology of intelligent information technology, analyze the main sources and methods of information risk detection, and hope to provide clues for the comprehensive application of big data and artificial intelligence technology [1]. Authors such as Huang Hong, Xiong Zhuo, Wang Yu, Lei Taoling, Wang Junqi, etc. stated that the basic positioning of big data technology is a multi-dimensional projection and understanding analysis system. The program completes the previous process in the landscape data field by creating a multi-dimensional view system for parallel applications and creating a single IoT agent, thereby identifying multi-dimensional, real-time and panoramic views of the application. The platform integrates various data sources of the most advanced cloud data platform, provides in-depth understanding of application data and diagnostic data, and ensures accurate application maintenance [2]. Applying big data intelligence technology to various fields can improve the intelligence of this field.

This paper decides to study the financial software perception function of intelligent information technology based on data mining as the research direction. Among them, Among them, the decision tree algorithm used in intelligent information technology, The Logistic algorithm and decision tree algorithm in data processing technology are applied to the data prediction designed by the perception function of financial software, the prediction results of the two algorithms are compared, and the accuracy of the data obtained exceeds 80%. Use the specific data content of the experiment in this article to prove the feasibility and authenticity of the method proposed in this article [3].

2 Overview of Big Data and Financial System

2.1 The Current State of Big Data

In today’s world, globalization, networking, and digitization are becoming more and more common. The speed of the Internet is gradually accelerating, technology is changing with each passing day, and terminal applications are diversified. Everyone’s life is inseparable from data transmission, search, storage and application. Network data has also started the trend of explosive growth. The measurement category of network data has also increased from basic bytes (B) and kilobytes (KB) to megabytes (M) and gigabytes (G). The amount of data we usually use is in this Russian-Japanese unit., And then to trillion bytes (TB), the data volume will reach the current one hundred trillion bytes until it exceeds the petabytes, tens of billions, etc., according to this growth rate, the network data volume will be Close to 40ZB. Experts say that such a large amount of data allows people to analyze data in social sciences and daily life, and draw accurate conclusions and judgments based on this [4].

2.2 Narrow and Broad Big Data

The characteristics of intelligent information technology generally recognized in the current academic circle are summarized as 7 V, namely Volume, Speed, Value, Variation, Truth, Visualization and Validity. The specific descriptions are as follows [5]:

- (1) In terms of speed, it depends on the amount of captured data, your storage, search, analysis and utilization must also be fast. Otherwise, you can only be submerged in the large amount of data captured.
- (2) In terms of value, because the total amount of data is fixed or limited, Due to the massive increase in data, the value of your single data decreases, and the amount of unit value gradually decreases, so how to mine useful information is a difficult problem faced by big data.
- (3) In terms of types, because big data technology is filled with different data in many places, the data information has been replaced from the single structure of the data stored in the past to the current unstructured data. Now there are sound, video, location, and real-time tracking. Wait for the data, the latter is difficult to operate before, what development trend has been shown.
- (4) In terms of accuracy, because intelligent information technology emphasizes object compatibility, the relationship between things is often ignored. This leads to a decrease in the accuracy of the data.
- (5) In terms of visualization, big data technology is trying to visualize the current data as much as possible, that is, more pictures, tags, sound effects, videos, etc., reduce the margin of data analysis, and make more changes. Have a professional background People can mine, analyze and process non-professional data more flexibly, and that's why. Promotion and development of streaming media technology.
- (6) On the formal side, Although the development of intelligent information technology has only been in recent years, its role, influence and changes are far-reaching, and it has brought a series of behavioral, legal, technical, political and social consequences. Therefore, safeguarding and promoting one's own rights is an issue that must be paid attention to in the future.
- (7) In terms of capacity, the data capacity of big data can be called a huge amount, which has been measured in tens of billions of bytes (ZB) [6].

In a broader context, big data technology should also include technical methods and thinking models for the operation, storage and analysis of big data technology, and further include talents, teams, company services, governments and companies dealing with big data. In addition, for "big data", some scholars interpret it as "verbs", which is another way of expressing and narrating "analysis and processing". In this sense, it can also be understood as the storage, mining, analysis, processing, And broad data applications. During this time, intelligent information technology has become a way to compile and process data in exchange for valuable decisions, rules, research judgments, predictions, etc. [7].

2.3 Application of Data Mining

(1) Logistic regression algorithm.

Both the traditional linear regression model and the general linear regression model require the dependent variable to be a normally distributed variable, and there is a linear relationship between the independent variable and the dependent variable [8]. Logistic regression is an important regression model that can also be used for prediction. Mainly used for binary isolation analysis. Unlike traditional regression models, the response variable is classified rather than progressive [9].

The principle of the algorithm is as follows: Suppose there are samples $\{x,y\}$, y is 0 or 1, which is used to represent the positive class or the negative class, and x is a sample feature vector with m dimensions. Then for the sample x belongs to the positive class, that is, the probability of $y = 1$ can be expressed by the formula function:

$$p(y = 1|x:0) = \sigma(\theta^T x) = \frac{1}{1 + \exp(-\theta^T x)} \tag{1}$$

Among them, θ is the data parameter of the model, which is the regression coefficient, and σ is the function of sigmoid, which can be transformed into:

$$\log it(x) = \ln \frac{p(y = 1|x)}{p(y = 0|x)} = \ln \frac{p(y = 1|x)}{1 - p(y = 1|x)} = \theta_1 + \theta_2 x_1 + \dots + \theta_m x_m \tag{2}$$

That is, y as a dependent variable is related to multiple independent variables. Here, these independent variables are expressed as x_1 to x_m corresponding to the weight, and θ_1 to θ_m correspond to the regression coefficient of the independent variable.

(2) Decision tree algorithm.

Decision tree is a very common technology that can be used for data analysis and classification, and it can also be used for prediction. The machine learning technique for generating decision trees from data sets is called decision tree learning, which is a technique for generating decision tree models based on existing data and using the models to predict the future of classification. The advantages of decision tree are high classification accuracy, simple generated model and good fault tolerance to noise [10].

The basic idea of ID3 algorithm is to select a feature with judgment ability attributes as the node of the current decision tree analysis according to information changes. The algorithm principle is as follows. Suppose S is a sample set containing s data, the attribute has m different classes c_i , and s_i is the number of samples contained in class c_i , then the calculation formula for information entropy is:

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i) \tag{3}$$

The p_i in the formula is the probability of any sample c_i .

2.4 The Perception Function of the Financial System

Today when intelligent information technology is popularized, financial management has the characteristics of computer-aided, and financial management is a feature of computer-aided computing. In information technology, computer software has the characteristics of calculating complex data. Need to use new software technology to improve the efficiency of computer systems and employees. In terms of financial software, we use enough data to predict the next operation of the service, provide users with more efficient and effective strategies, and greatly improve the intelligence of financial software [11].

In this article, we will combine our understanding of the intelligent perception function of financial software to propose a more efficient data mining algorithm. Analyze the application rules according to the operational knowledge level. When using software in data mining, there must be a lot of information, and this information is hidden in a lot of data, So use the algorithm in the software to extract the associated data in the software. Effectively improve the security and intelligence level of financial software. In any case, the extraction of good behavior and action data can be done in the user interface. Under normal circumstances, data will be displayed in the form of digital codes in the software of the financial system, but the randomness and uncertainty of digital codes will cause financial managers to waste a lot of time and reduce resource utilization [12].

3 The Financial Software Perception Function of Big Data

3.1 Technology and Software Used in the Perception Function of Financial Software

- (1) J2EE is essentially a server application server environment and a Java environment, including: the central application operating system environment and the Java extended API set for building applications can be used to create an enterprise-level distribution with features such as scale, flexibility, and flexible maintenance application. By providing a unified development platform, reducing the cost and complexity of developing multiple systems, while integrating existing facilities, with better performance and safety systems, the ultimate goal is to become a large-scale architecture to shorten the development cycle of the industrial system.
- (2) Myeclipse is an enterprise-level development application that integrates a large number of development-based plug-ins and development tools. The platform adopts a modular design, which can be expanded and upgraded at will without affecting other units. Myeclipse can be used as a powerful J2EE development environment that supports coding, configuration, testing, debugging and application verification.

By adopting the B/S architecture and Maven configuration process, you can easily manage and change configuration settings. Developed based on java technology, easy to move and replace the platform, mature technology, more related documents and applications. The program has a complete log mechanism, which can easily confirm the operation of the system and facilitate the troubleshooting of errors and exceptions. First, we must analyze and classify the service data area, identify the relevant subject areas, and then model the database in the relevant subject areas. In the process of model design, we must

think about the impact of the model on the selected database, and make full use of the characteristics of the database to improve the credit efficiency of the model.

3.2 Experimental Process

On the Internet, the data transaction volume published by a certain securities trading website is used as the basis, the data transaction volume of the previous year is used as the independent variable, and the transaction data volume of the following year is used as the dependent variable. Substituting the data of the previous year as the base number into the Logistic regression algorithm and decision tree algorithm proposed in this article to calculate and predict. After the calculation, compare the calculated data with the data of the next year and calculate the mold to get these two calculation algorithms. The accuracy rate.

4 Big Data Intelligent Technology Financial Software Perception Function Experiment

4.1 Logistic Algorithm and Perception Test

This paper uses the Logistic algorithm and the decision tree algorithm to analyze, and then integrates it into the perception function of financial software, and compares the information obtained by the two algorithms, so as to achieve the result on the basis of integrating the two algorithms into the perception function. The corresponding accuracy rate. Take the data in the two-year financial report on the securities website on the Internet in a certain year as variables, conduct data practice, and give data conclusions. Specific data shown in Table 1.

Table 1. Logistic algorithm regression results table

Parameter	Parameter estimates	Standard deviation	Wald statistics
Past asset income	- 13892	3872	12818
Past net profit growth	- 5292	1465	10287
Sales growth rate	- 3290	0.90	5233

As shown in Fig. 1, the ground difference indicators on this website are all negative numbers, and then according to the Logistic regression algorithm, the data can be substituted into the formula to get the intercept of the parameter estimate as 2593.

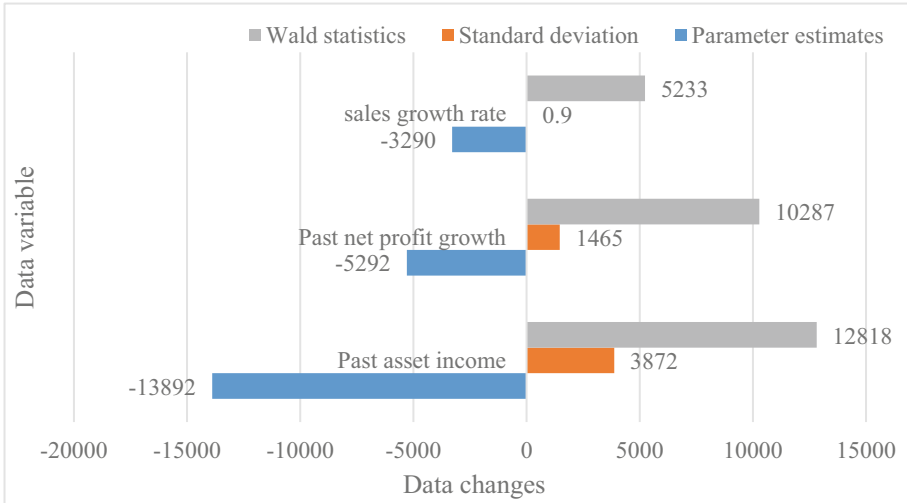


Fig. 1. Logistic algorithm regression calculation diagram

4.2 Prediction Experiment of Logistic Regression Algorithm and ID3 Decision Tree Algorithm

As shown in Table 2, the accuracy of the prediction results of the two algorithms exceeds 80%, which shows that it is feasible to apply the data mining technology in big data technology to the perception function of financial software. The perception software is also relatively accurate in predicting the data trend of the future market.

Table 2. Predictive analysis results table

Algorithm	Roe	Forecast result		Accuracy
		More than 15%	Less than 15%	
Logistic	More than 15%	261	53	80.80%
	Less than 15%	252	1030	
Decision tree	More than 15%	296	18	86.37%
	Less than 15%	198	1084	

As shown in Fig. 2, the prediction algorithm using the decision tree is better than the Logistic regression algorithm. Both algorithms can give the corresponding data trend of the company’s assets in the future, and whether it is an increase or decrease in assets, both algorithms have a higher accuracy rate, which improves the accuracy of financial software analysis.

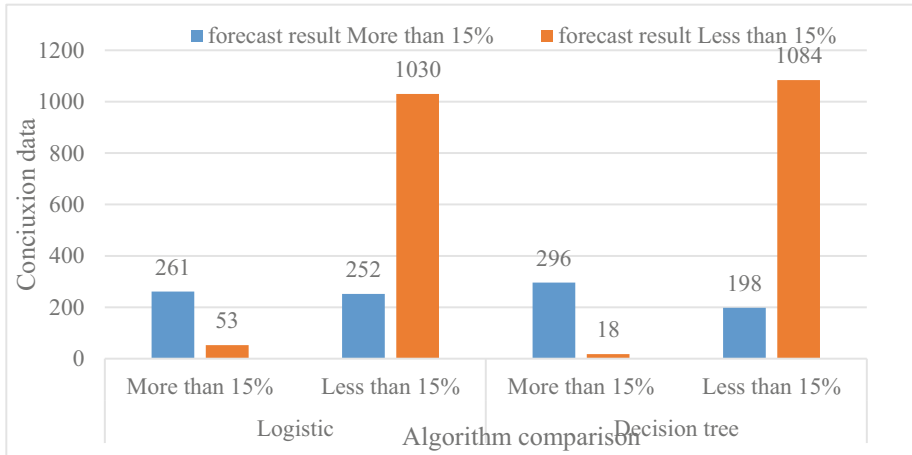


Fig. 2. Two algorithms predict data graph

5 Conclusions

This article introduces the concept of intelligent information technology in big data in detail, and then applies the data mining technology in big data intelligent information technology to the design of financial software, and adds a new data perception function to the financial software. This function can predict the future trend of data flow based on the analysis of the corresponding data. The research direction of the perception function of financial software is introduced and the algorithm model is given, and then data experiments are carried out through the algorithm and the data obtained on the network, and the experimental results are given.

References

1. Mohammed, S., Arabnia, H.R., Xiaobo, Q., Zhang, D., Kim, T.-H., Zhao, J.: IEEE access special section editorial: big data technology and applications in intelligent transportation. *IEEE Access* **8**, 201331–201344 (2020)
2. Roitberg, A.: *Uncertainty-aware Models for Deep Learning-based Human Activity Recognition and Applications in Intelligent Vehicles*. Karlsruhe Institute of Technology, Germany (2021)
3. Zhang, X., Wang, Y.: Research on intelligent medical big data system based on Hadoop and blockchain. *EURASIP J. Wirel. Commun. Netw.* **2021**(1), 1–21 (2021)
4. Salman, O.H., Taha, Z.K., Alsabah, M.Q., Hussein, Y.S., Mohamed, A.S., Aal-Nouman, M.I.: A review on utilizing machine learning technology in the fields of electronic emergency triage and patient priority systems in telemedicine: Coherent taxonomy, motivations, open research challenges and recommendations for intelligent future work. *Comput. Methods Programs Biomed.* **209**, 106357 (2021)
5. Wan, J.C., Cheng, T.: Research on multi-agent and storm-hadoop based on cooperative sensing framework for multiple intelligent manufacturing agent. *Appl. Mech. Mater.* **864**, 192–201 (2017)

6. Polap, D., Srivastava, G., Yu, K.: Agent architecture of an intelligent medical system based on federated learning and blockchain technology. *J. Inf. Secur. Appl.* **58**, 102748 (2021)
7. Joyo, A., Yaqub, R., Madamopoulos, N.: Intelligent traffic-lights management by exploiting smart antenna technology (ITSAT). *IEEE Intell. Transp. Syst. Mag.* **13**(4), 154–163 (2021)
8. Poonguzhali, S.: Rekha Chakravarthi: A sensor based intelligent system for classification and assistance of diabetes patients in telemedicine technology. *J. Intell. Fuzzy Syst.* **40**(4), 6365–6374 (2021)
9. Xie, J.: Research on Weibo user behavior system for subjective perception and big data mining technology. *J. Intell. Fuzzy Syst.* **38**(2), 1–10 (2019)
10. Chen, S.: Analysis of Customization Strategy for E-Commerce Operation Based on Big Data. *Wirel. Commun. Mob. Comput.* **2021**(6), 1–11 (2021)
11. Shin, W.S.: The perception types of the directors on PCC financial management and its improvements. *J. Digi. Convergence* **15**(2), 1–9 (2017)
12. Lu, C., Liu, J., Liu, Y., Liu, Y.: Intelligent construction technology of railway engineering in China. *Front. Eng. Manag.* **6**(4), 503–516 (2019)



Effective Combination of Intelligent Information Technology and Labor Education Data Analysis Algorithms

Shasha Lv^(✉)

Zibo Teachers Technical College, Zibo 255130, Shandong, China

Abstract. With the continuous deepening of curriculum reform, the examination-oriented curriculum has evolved into quality education, a new teaching model that emphasizes both theoretical knowledge and practical skills, and has become the consciousness of the society and education community. As a subject of cultural education, the labor education system has gradually changed due to the influence of schools, students and parents. However, the traditional methods of labor education and training are the same and may not meet the requirements of curriculum development in the new era. Therefore, research should be carried out on the combination of intelligent information technology and labor education based on functional characteristics. The purpose of this article is to study the effective combination of intelligent information technology and labor education. This paper determines the effective combination of intelligent information technology and labor education as the research goal, and then determines the data analysis algorithm based on the reference case and the data obtained, and compares the data obtained in the practice of this paper with the reference case. Based on the research on intelligent information technology and labor education, based on a large number of surveys and data comparisons, this paper comprehensively considers the advantages and disadvantages of the research direction, and considers the specific implementation of the plan from the perspectives of technical level, teachers and students factor. Then this paper uses data experiments to prove the validity and authenticity of the research direction and purpose of this paper. According to the comparative experimental data, based on the S-analysis method proposed in this paper and the calculation using the algorithm therein, the overall score and the passing rate of the examination in the experimental class assisted by intelligent information technology are higher than those in the reference class using traditional practical teaching. The average score is about 5.4 higher, and the passing rate of the assessment is 4.4% higher. The data has improved significantly.

Keywords: Intelligent information technology · Labor education · Effective integration · School education

1 Introduction

Since the rapid rise of “multimedia technology” in information technology and Internet technology in the 1990s, it has swept the world. Since the beginning of the new century, in

the world, intelligent information technology has developed rapidly in many areas of our country with its own unique development method, improving all aspects of people's lives. Labor education refers to the use of specific methods by schools or other educational institutions to teach students a lot of knowledge and labor skills about labor, and to understand labor skills to a certain extent.

In the research based on the effective combination of intelligent information technology and labor education, many experts, scholars, and teachers on the front line of teaching have conducted research on it and achieved good results. For example, Li Hongfei once said that in the new round in the basic education curriculum reform, the comprehensive practical activity curriculum has attracted more and more attention. Labor and technical education is to allow students to engage in operational learning through the role of people and things, the interaction between people, and the combination of hands and brains. So as to obtain comprehensive development [1]. Jiang Rong also pointed out that labor education is an important part of quality education, and it is the key to coordinating moral, intellectual, physical, and aesthetic education. With the rapid development of information technology, information technology education in higher education will pay more attention to the cultivation of students' information literacy and computational thinking ability. This requires attention to the process of practical exploration in the information technology courses of higher education, which is to teach through fun and education. Labor, organically integrate information technology education with labor education [2]. Labor education is an indispensable important part of research and teaching in our country's teaching.

This paper determines the research goals of the effective combination of intelligent information technology and labor education, ST analysis methods, and the parameters of data experiments, and then uses examples of this paper to verify the operability and effectiveness of the research directions and goals, and then uses a variety of The method was compared to prove the verifiability of the experiment in this paper.

2 Intelligent Information Technology and Labor Education

2.1 The Significance of Labor Education

Labor is an important means for human survival and development, as well as an important way to improve the quality of one's own will and realize the value of life. With the rapid development of social economy, different values and diversified social thinking are presented. In such an environment, labor education is facing more opportunities and challenges. How to find practical and effective methods and means to optimize students' labor education curriculum, how to better understand labor values, labor ethics, and labor spirit to develop correct cognition, so as to better help students develop, it is Problems that need to be solved in the current labor education research [3].

The factors that affect the labor education material rate are the student's family, school choice and social attention. The labor education mentioned by scholars refers to the labor education of school education. Scholars have basically the same views on labor education, but the reference objects are different.

Some scholars believe that the labor education program is a skill education program whose main purpose is to enable students to master labor skills in the process of

promoting labor emotional performance and overall quality improvement through labor programs.

Other scholars pay more attention to the function and practicality of labor training, and believe that the purpose of labor training is to gain more academic understanding through labor experience, and to better understand society and oneself through labor training [4].

2.2 The Necessity and Purpose of Student Labor Education

Labor education is an important part of student quality education. It is an important part of the student's education plan. Educational content is very important to promote the all-round development of college students. It is an important link in creating a good social environment and meeting social needs in line with the national education policy.

The purpose of labor education is to cultivate students' appropriate labor values, enhance labor emotions, develop labor habits, master labor skills, cultivate labor discipline, promote the development of labor ability, promote youth labor awareness and labor quality, establish a positive attitude, and cultivate The diligent spirit of students promotes the common development of labor education and school behavior education, spiritual education, virtue and sports [5].

2.3 Methods of Carrying Out Student Labor Education

(1) Raise the necessity of students' awareness of labor education

Students are the first part of labor training. Giving full play to their creative thinking is the core and key to effective labor education programs. Require students to actively participate in labor training, the most important thing is to raise students' awareness of the importance and necessity of labor education [6].

As today's society has higher and higher requirements for the quality of talents, and the current work situation is becoming more and more severe, students will face many challenges such as personal growth and employment. If you do not improve your own labor quality, it will be difficult to adapt to the needs of social development. What modern society needs is talents with morals and talents. They must have strong knowledge and skills, and they must also have good behavior and good knowledge. Therefore, students and students must pay full attention to the labor curriculum and cultivate their own independent ability and ability. Good sense of responsibility, cultivate the correct value of work, improve the quality of students, make contributions to their own future, national development, and social progress, and lay a good foundation [7].

First of all, students need to set reasonable values in their lives. People's thoughts and behaviors are affected by their values. Students should establish lofty values in life, knowing that the realization of their own value is not in personal enjoyment, but in their contribution to the country, to society, and to others. Only by establishing the correct values of life and being aware of the responsibilities of oneself, can the self-consciousness to improve one's character and ability be produced [8].

Secondly, students must clearly recognize their own problems. While establishing lofty values in life, students must clearly recognize their own problems and gaps. Students should clearly recognize their own problems and the gap between themselves and the requirements of society, and consciously seek changes. Only by doing well, can they avoid unnecessary setbacks when they are unprepared to join the society.

Finally, students must realize the necessity of labor education. To overcome their own problems, student labor education is indispensable. Obtain correct labor values through labor education, gain independence in life and personality through labor practice, enhance the psychological quality in the face of setbacks, and enhance their own morality and ability. A correct understanding of the necessity and importance of labor education is a prerequisite for fully mobilizing students' enthusiasm for participating in labor education and triggering students' consciousness to continuously improve their comprehensive quality through labor practice [9].

(2) Raise the school's awareness of labor education

In the new historical period, colleges and universities need to increase the importance of labor education to student education. This is not only the realization of the national teaching plan, but also the need for actual education. The development of modern society is inseparable from talents, and the demand for talents is getting higher and higher. Higher education is the beginning of cultivating talents for the society. What kind of talents a university can teach is directly related to the development and construction of the country. Universities are responsible for creating high-quality new talents for the country and society. Need to keep pace with the times to change traditional concepts. They need to understand the current situation of students and understand the requirements of the society for talents in order to find the correct and effective way to cultivate talents.

First of all, in the classroom, professional teachers of colleges and universities can provide students with labor education in combination with students' learning content. On the one hand, it can improve students' appreciation of works, and on the other hand, it can make students feel the charm of labor education in the course. Improve the realm of labor education, and taste the benefits that labor education can bring to students from a philosophical point of view [10].

(3) Strengthen the construction of campus labor education

The campus division can have a big impact on the three views of students, which is unparalleled by any experience. A good culture can have a positive effect on improving student behavior, behavior training and healthy behavior. At present, most students live in the school. During the precious four to five years of university time, it has a positive impact on the cultural construction of labor education in colleges and universities, and it has a positive impact on students. This is an important part of gardening culture [11].

(4) Carry out rich labor education activities

The first is to carry out lectures, seminars, seminars and other activities related to labor training. Local staff or professional team staff can be invited to the school to teach on-the-job teaching, and the teacher will personally guide teachers and students to complete the homework. A discussion competition related to labor education is held every semester to increase students' emphasis on labor education. The second is to perform collective tasks related to labor education. From the perspective of individual trainees, the on-the-job training team is also a closely integrated physical training project. Dormitory activities are labor education activities that take multiple dormitories as a unit or the entire dormitory as a unit, and can be motivated by dormitory labor appraisals. The third is to use campus media for education. Make full use of the highly developed media in the new era, such as school post bars, official accounts, and other multimedia software used by students today. Publish the content of labor-related course topics so that students can view and respond on the Internet during the holidays [12].

The forms of labor education activities are diversified, and students will gain a lot in the process of participation, but in the process of carrying out activities, they need the correct guidance of educators to truly play the role of education in labor education.

3 Data Practice of Intelligent Information Technology

3.1 The Purpose of the Experiment

The combination of intelligent information technology and work training fully embodies the practicality, independence and creativity of practice. Use information technology to create emotional or problematic situations, encourage students to think, let students freely use the resources provided by the Internet for practical and research training, deepen their understanding and understanding of the content of the course, and let students choose to collaborate around key themes and participate in teachers and teachers. Multi-dimensional dialogue between students, student books and students, and conduct joint research. Then teachers and students collect information together, select and process information, form consensus or ask more detailed questions.

3.2 Experimental Comparative Analysis

It is the product of the combination of intelligent information technology and labor education. Through this intelligent information technology, students can realize the joy of practice and immerse students in the practice of labor courses. Then combined with the actual manual operation.

For example, in the electrician training selected, the ST analysis method was used to compare and analyze the two different teaching methods of the experimental group and the reference group with the same learning content. It needs to be explained that the selected project focuses on the mastery of professional basic knowledge and belongs to practice Experimental teaching in the course. Guide students to use information technology to complete the research of each topic through group cooperation, so as to complete

the learning task. Jiang has one class as an experimental group, which combines intelligent information technology for data practice, and the other class as a reference class, using traditional practical teaching methods. After a month of teaching to participate in the assessment, the teacher will evaluate the scores. Each class is divided into 10 groups and an average of 5 people in each group. The statistics of the scores are evaluated according to the credentials of each group.

3.3 Principles of S-T Algorithm

T’s analysis method is based on ST. At the enlightenment moment on the way, the actual detected S behavior and T behavior time are displayed on the S axis and T axis in the order of acquisition according to the measured S and T verses until the data The end of the acquisition.

Rt represents the occupancy rate of T-lines, that is, the proportion of T-lines in the teaching process:

$$R_t = N_T/N, \tag{1}$$

Where N is the total number of samples, and N_T : T is the number of rows.

Ch represents the conversion rate, that is, the ratio of the number of conversions between T rows and S rows to the total number of rows. The specific formula is:

$$Ch = (g - 1)/N \tag{2}$$

Among them, g is a losing streak, and the same behavior is a continuous losing streak.

4 Practice Control Experiment Based on Intelligent Information Technology

4.1 Reference Class Assessment Data

As shown in Table 1, in the traditional practical teaching, the scores of the students in the reference group are basically concentrated between 60 and 70 points. The average score of the entire reference book is around 65.5. The pass rate of the assessment was 90.0%. Among the 50 students in the reference class, 11 had a score of 70 or higher.

Table 1. Evaluation data of the reference class

Group	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
The average score	70	57	74	63	64	61	63	60	70	73

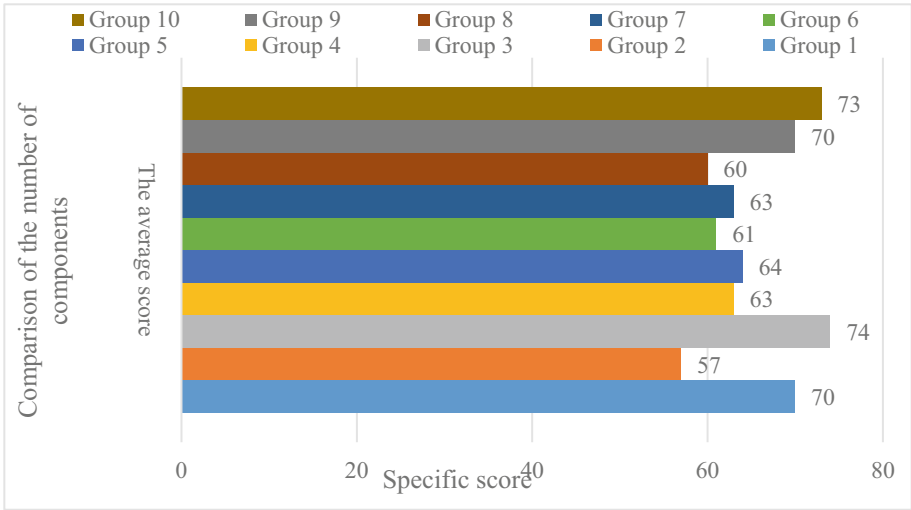


Fig. 1. The average grade chart of the reference class

As shown in Fig. 1, from the group performance distribution curve, 11 students in the reference class have a score of 70 or higher, and 6 students with a score of 22% or less account for 12% of the total number of students. There are three groups with an average score of 70 or more.

4.2 Evaluation Data of Experimental Class

As shown in Table 2, looking at the average scores of the 10 groups in the experimental class, most of the students' scores are concentrated between 65 and 80, the pass rate of the assessment is 94.4%, and the average score of the entire class is 70.9. Among the 50 students in the experimental class, 23 had a score of 70 or more.

Table 2. Examination score table of experimental class

Group	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
The average score	80	68	74	77	65	65	62	70	80	68

As shown in Fig. 2, both the average grade and the pass rate of the classes that use intelligent information technology to participate in practical teaching exceeded the reference class. From the results, the effect of using intelligent information technology combined with labor education is higher than the traditional teaching method combining teaching and practice.

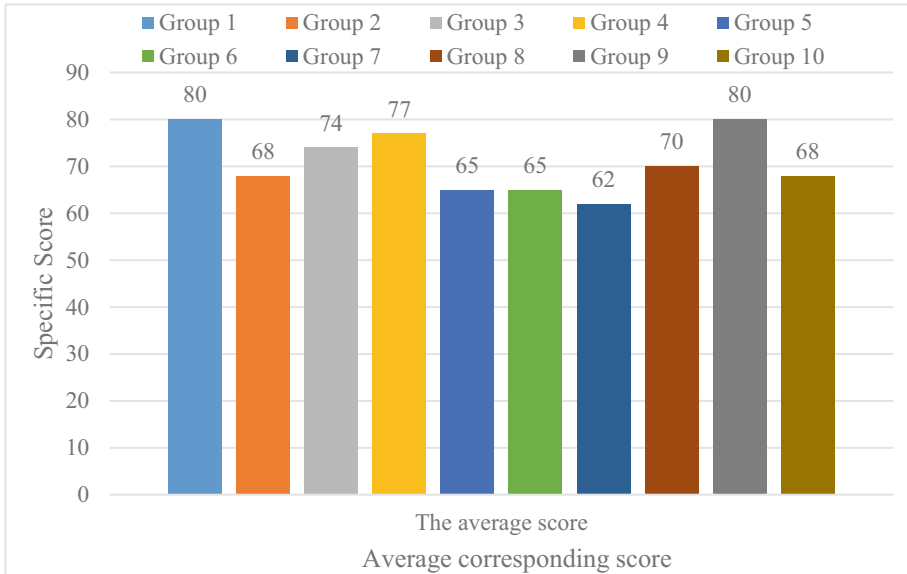


Fig. 2. The average score of the experimental class

5 Conclusions

This article combines intelligent information technology with labor education, focusing on school students' emphasis on labor, supplemented by intelligent information, and setting up lofty life values to illustrate the importance and necessity of labor education for our students. Emphasizes the importance of labor education to the training of talents in our country. Then analyze the loopholes and shortcomings of China's high-efficiency labor education, and give specific implementation methods for cultivating high-quality and high-quality new-type social talents.

References

1. Li, H.: The development of comprehensive practical activity courses with "labor and technical education" as the starting point-Dalian Ganjingzi District Student Labor Skills Training Center. *Urban Family Edu. Monthly* **8**, 271–272 (2017)
2. Jiang, R.: Analysis of the relationship between information technology courses and labor education in higher education. *2021(2020–6)*:149–150
3. Piyavsky, S.A., Kiryukov, S.R., Kuznetsov, A.S., et al.: Information technology for career guidance of creatively gifted university students. *Inform. Educ.* **8**, 6–15 (2020)
4. Zhang, L., Yuan, H., Chang, S.-H., Lam, A.: Research on the overall architecture of Internet of Things middleware for intelligent industrial parks. *Int. J. Adv. Manuf. Technol.* **107**(3–4), 1081–1089 (2019). <https://doi.org/10.1007/s00170-019-04310-z>
5. Zhao, Z.: Research on the evaluation of public policy execution ability with picture fuzzy information. *J. Intell. Fuzz. Syst.* **37**(2), 1659–1667 (2019)
6. Javier, R.: Integration of service robots in the smart home. *Intell. Syst. Control Autom. Sci. Eng.* **53**, 115–142 (2017)

7. Marinov, R.: New Intelligent Technologies—Interactivity and Information Issues. *Adv. Journal. Commun.* **07**(3), 94–108 (2019)
8. Ahmed, D.: Artificial intelligence and contemporary Japanese architecture – any relationship? *Intell. Build. Int.* **12**(4), 1–14 (2019)
9. GUO, Qian, LI, et al. Evaluation of intelligent tourism information service capabilities: A case study of Hanzhong City. *2021(2017–1):2–13*
10. Enrique, D.V., Druczkoski, J., Lima, T.M., et al.: Advantages and difficulties of implementing Industry 4.0 technologies for labor flexibility. *Proced. Comp. Sci.* **181**(19), 347–352 (2021)
11. Noh, S.K.: Recycled Clothing Classification System Using Intelligent IoT and Deep Learning with AlexNet. *Comput. Intell. Neurosci.* **2021**(4), 1–8 (2021)
12. Zavadzka, N., Ignatiuk, M.: Forming of Information Competence as a Precondition of Academic Integrity: the Role of the Library of Higher Education Institution. *Ukrainian J. Libr. Inf. Sci.* **5**, 44–65 (2020)



Chinese Language and Literature Relational Database Mining Based on Association Rules Algorithm

Yingxia Tang^(✉)

Radio and TV University of Jilin City, Jilin City 132000, Jilin, China

Abstract. With the ever-increasing demand for high-quality and high-quality teachers in society, the professional certification of teacher training aimed at ensuring the professionalism of teacher education has been fully implemented in our country. This article mainly studies the application of association rule algorithm in the reform of Chinese language and literature majors in colleges and universities. The collection of student performance data information is a process of collecting information such as student ID, name, academic performance of each subject, and learning evaluation. Data extraction is to select the data needed in our mining tasks from various relational databases of our school, integrate it into the data warehouse, and prepare the data for our future mining. This article selects some courses closely related to Chinese language and literature for excavation, trying to find out the relevance of these courses and the support of these courses for obtaining vocational qualification certificates, so as to provide us with theories for strengthening professional skills training, professional construction and curriculum reform in accordance with. The scores of general education courses in the five normal universities fluctuate between 23.84% and 31.23%, and the average score is 28.4%. The results show that the association rule algorithm provides theoretical support and practical guidance for the reform of Chinese language and literature majors in colleges and universities.

Keywords: Association rule algorithm · College Chinese language and literature · Curriculum reform · Relational database

1 Introduction

With the reform of basic education curriculum, it has attracted great attention all over the world, especially in some developed countries, education reform generally starts from the reform of basic education curriculum. Through the reform of basic education curriculum, the adjustment of talent training objectives, the change of talent training methods and the improvement of talent training quality, all countries put the reform of basic education curriculum first.

With the rapid development of science and technology, the process of economic globalization is accelerated, and the competition of comprehensive national strength is becoming more and more intense [1]. In order to meet the requirements of the era of

talent training, countries all over the world have put the reform of higher education in a very important strategic position [2]. Therefore, it is very important to adjust the training objectives and curriculum of the University in time [3]. In the educational administration system of the University, only one structure is used. Both C/S structure and B/S structure have some disadvantages [4]. Therefore, according to the characteristics of the system, the advantages of the two structures can be used to build a hybrid structure that can be used with each other [5]. Therefore, it can flexibly complement each other and give full play to the two more beneficial structural advantages in the educational administration system [6]. In the case of fierce competition, although the theoretical learning time is shortened, the training methods must not be reduced. On the contrary, we must strengthen the training of vocational skills and emphasize the strength of vocational trainees [7, 8]. In order to adapt to the changes of the new research mode, the most urgent task is to reform the existing curriculum system and realize the talent training standard in less teaching time [9]. When students study, they must start their research in a specific order. For example, before learning professional courses, you first need to learn the basic courses of professional knowledge. Moreover, the learning effect of previous courses has a great impact on the learning effect of subsequent courses [10].

In the curriculum reform of colleges and universities, data mining technology is used to mine the useful information accumulated by students in the performance management system. We can adjust the curriculum to the relevant departments of the school, promote the adjustment of the time number of popular courses and unpopular courses, and encourage students to select complementary courses and give guidance by selecting courses and relevant information. This will not only help to improve the overall quality of students, but also help to reasonably allocate educational resources. The relevant rule algorithm is segmented and changed, and the two steps of Apriori algorithm are clearly segmented. Frequent itemsets will be sorted and read by the distributed system in order. Then, parallel computation is repeated to obtain frequent item sets. As a result, association rules are generated using the MapReduce framework. Finally, the association rule algorithm can be successfully applied to the cloud computing distributed platform, and the distributed computing function of Hadoop platform can be used to realize the parallelization of Apriori algorithm operation and memory computing, which can reduce the disk I / O overhead and realize the high-speed reading, writing performance and application efficiency of Apriori algorithm.

2 Curriculum Reform of Chinese Language and Literature Majors in Colleges and Universities

2.1 Association Rule Algorithm

In this process, the core of association rule mining is to find all frequently used itemsets. When using association rule mining, you need to repeatedly scan the database D as the first solution, and it takes a lot of time to repeatedly scan the database and perform input and output operations. Therefore, the main problem solved by various related rule algorithms is to find all frequent item sets quickly and effectively. This is also the standard for measuring various related rule algorithms. Each partition needs to be

placed on the computer's main memory for processing, and each partition needs to be scanned once. Therefore, the computer usage must be fully considered when creating the partition. The correctness of this algorithm is because the frequent itemsets of each possible block may occur frequently throughout the block. Because the algorithm is processed individually in blocks, the algorithm can be processed in parallel. It can be done using multiple processors or distributed operations. However, the communication process of each block and the time required for each independent processor to generate the frequency set are the bottleneck of mining efficiency.

With the rapid development of computer technology and the widely used database system, more and more data are stored in the database. Under the background of the surge of data information, there is unknown but very valuable information. It is hoped that through careful analysis and investigation, we can explore these useful and important information. The database system performs data input, query and other businesses, but it cannot make full use of data and information, find potential rules from multiple data sets, or predict the development trend of things. Because there is no effective means to mine the information hidden behind the data, all kinds of data will increase explosively, but the knowledge is insufficient. This strange phenomenon aroused people's thinking. After continuous research, data mining technology came into being. The so-called data mining is that it is difficult to find valuable information and knowledge from a large amount of data. These hidden information is useful knowledge for people. The mined knowledge is expressed in the form of concepts, models, laws, rules and so on.

The model inputs H into the Softmax classifier to obtain the parsing result u_t of each sentence s_t , and gives the conditional distribution probability on the specific task category:

$$u_t = \sigma(W_y H_t + b_y) \quad (1)$$

$$p(y_t = c | H_t) = \frac{\exp(e^{u_c})}{\sum_{j=1}^{|C|} (e^{u_j})} \quad (2)$$

Among them, (W_y, b_y) is the calculation parameter and bias term of the classifier. The formal definition of the negative log likelihood function is as follows:

$$J(\theta_e, \theta_c, I) = - \sum_{t=1}^N \log P(y_t^i | \theta_e, \theta_c) \quad (3)$$

2.2 Curriculum Reform of Chinese Language and Literature Majors in Colleges and Universities

The knowledge points recognized for the professional qualification of Chinese and literature in universities are included in each course, and the proportion of each course is different, which is directly reflected in the performance of the students in each professional course and the professional qualifications of the students. By analyzing the correlation between the performance of each professional course and the result of vocational qualification certification, we can find the support of each professional course

for vocational qualification certification, so the degree of relevance of the certification results is different. By teaching courses that strongly support accreditation, the pass rate of vocational qualification accreditation can be increased. Therefore, the combination of course content and students' actual life can shorten the distance between students and learning content, pay attention to the significance and value of learning content, and stimulate students. Concern for learning promotes students' independent and continuous exploration and improves students' creativity. However, the content of China's unique basic education curriculum ignores the problems of life and reality, and often focuses on knowledge and musical instruments. The content is profound and difficult, which is separated from students' life and social reality. Students' positive development, aiming at the problems of the previous curriculum content, the new curriculum reform emphasizes that students should learn more knowledge closely related to modern life. The main problems of Chinese traditional curriculum evaluation are as follows: the evaluation content unilaterally emphasizes knowledge learning and ignores the comprehensive test of practical innovation ability, learning methods and attitudes, psychological quality and value. Ignoring students' commonality and students' individual development; The evaluation method is mainly based on the single written test results, ignoring the students' learning process and progress; The main body of evaluation is teachers, and students participate in the evaluation of themselves and partners who fail. China's curriculum evaluation needs to focus on cultivating students, strengthen process evaluation and qualitative evaluation, and the evaluation objects must be diverse.

All colleges and universities should combine the actual situation of the school and the research group of Chinese literature education to cooperate in carrying out education, research and educational reform activities, so as to unite teachers, urge teachers to learn and make common progress. Only in this way can teachers of Chinese and literature education realize the reform of teaching concepts, such as their views on teaching plans, students, education and evaluation, and consciously carry out the reform of professional teaching plans. The goal is unbalanced, the unreasonable structure, the content separated from students' actual life, the teaching method can not emphasize students' subjectivity, and the evaluation of teaching plan is limited. Any reform is accompanied by the problem of interest distribution. The reason for the inertia of Chinese literature education teachers in the reform of teaching plan is undoubtedly related to the balance of interest. In the curriculum reform of this major, teachers need to spend more time and energy than previous education, so it is difficult to quantify many tasks and evaluate them. If there is no particularly important incentive, then the teachers of this major must have psychological inertia. Therefore, schools must formulate supporting policies and evaluation methods, work for the reform of professional teaching plans, and give professional teachers specific important remuneration. At the same time, because the performance of participating in curriculum reform is related to the promotion of professional teachers, professional teachers can obtain corresponding benefits, overcome inertia and improve enthusiasm.

3 Application Experiment of Association Rule Algorithm in the Reform of Chinese Language and Literature Major in Colleges and Universities

3.1 Data Collection

The collection of student performance data information is a process of collecting information such as student ID, name, academic performance of each subject, and learning evaluation. Data extraction is to select the data needed in our mining tasks from various relational databases of our school, integrate it into the data warehouse, and prepare the data for our future mining. For example, you can see in the student database: some students have no scores due to missing exams in certain subjects, and some students did not participate in the vocational qualification appraisal. Therefore, for students who do not have certain academic performances, the records of this part should be deleted; for students who have academic performances but have not participated in the vocational qualification appraisal, this part of the records should also be deleted. In short, the items with missing data should be deleted to avoid bias in subsequent mining.

3.2 Data Mining

This article selects some courses closely related to Chinese language and literature for excavation, trying to find out the relevance of these courses and the support of these courses for obtaining vocational qualification certificates, so as to provide us with theories for strengthening professional skills training, professional construction and curriculum reform in accordance with the program written in this design is implemented based on the classic Apriori algorithm, and the back-end database is SqlSever2005. This mining program is composed of two parts, namely data conversion and data mining.

4 Experimental Results

Figure 1 shows the overall proportion of teacher education courses in normal universities. The total course unit of the usual university is 165–180, the highest is K3, and the lowest is J1. The scores of the general education courses of the 5 ordinary universities fluctuate between 23.84% and 31.23%, and the average score is 28.4%. From the perspective of the credit ratio of professional courses, the credit ratio of the five ordinary universities fluctuates between 42.42% and 56.25%, and the total score averages 47.7%. From the point of view of the credit ratio of teacher training courses, the credits of the 5 schools moved between 16.81% and 22.67%, accounting for 18.6% of the total scores.

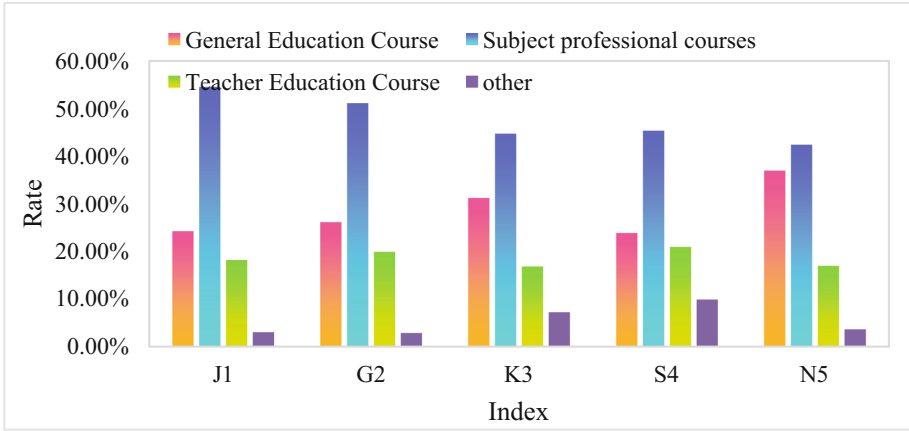


Fig. 1. The overall proportion of teacher education courses in normal universities

Table 1 shows the credit structure of compulsory and elective courses for teacher education courses in five normal colleges. There is little difference in the credit structure of elective and compulsory courses in teacher education courses of the five colleges and universities. The credits of required courses are all greater than 10, and the credits of compulsory courses in teacher education courses meet the requirements of teacher professional certification. The professional certification system requires teacher education curriculum content to introduce the latest results of basic education curriculum reform and education research, as well as excellent basic education teaching cases, and implement the cultivation of teacher ethics in the curriculum content.

Table 1. Credit structure of compulsory and elective courses of teacher education courses in five normal colleges

Colleges	J1	G2	K3	S4	N5
Compulsory course	22	22	20	29	20
Elective course	8	13	7	7	7

The results of the survey are shown in Fig. 2. When asked about “the content of teacher education curriculum is rich and practical”, the proportion of normal students who think that the content of teacher education curriculum is rich and practical is 38.7%, and the proportion of normal students who disagree with the content of teacher education curriculum is rich and practical is 35.2%. The proportion of students who choose normal teacher training courses is 26.0%. It can be seen that many normal students think that the teacher education curriculum of this major is not rich and practical enough.

This article analyzes the structure of relational databases, and analyzes the main characteristics of multi-valued, multi-dimensional, multi-layered and other related rules. The university courses in the new century pursue the value of “valuing people” and

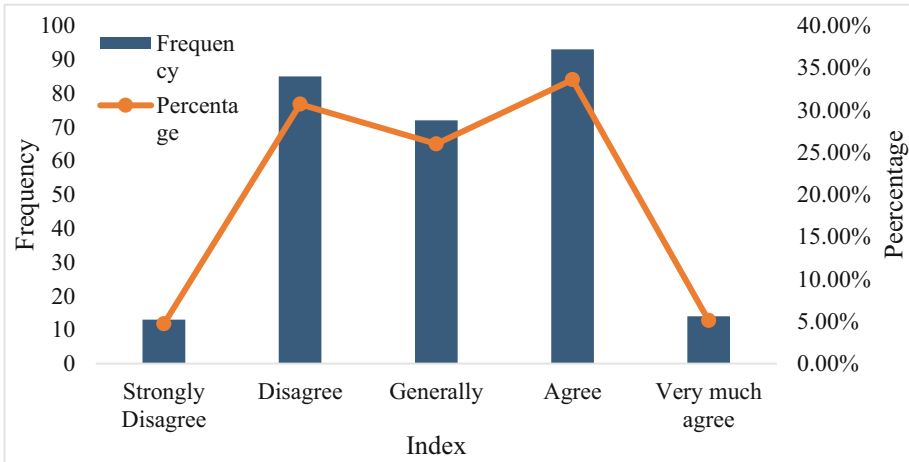


Fig. 2. Survey results

shift the focus to ideology and action. That is to say, just focus on knowledge, and it is necessary to realize the meaning of shifting the focus to existence and spiritual growth superior. In the curriculum reform, the characteristics of multidimensionality and openness are reflected, and the traditional single static and dull narrow curriculum must be changed. The research of data mining technology is an advantage. With the continuous expansion of demand and the deepening of research, a greater research climax will be formed in the future, and more results will be obtained. The DMQL language is gradually formalized and standardized, and the research on the visualization method of the data mining process has also made great progress.

References

1. Can, U., Alatas, B.: Automatic mining of quantitative association rules with gravitational search algorithm. *Int. J. Software Eng. Knowl. Eng.* **27**(3), 343–372 (2017)
2. Dbwan, B.A.: Enhanced K_way method in “APRIORI” Algorithm for Mining the Association Rules through embedding SQL commands. *INT. J. COMPUT. SCI. ENG.* **7**(10), 52–56 (2019)
3. Ait-Mlouk, A., Agouti, T., Gharnati, F.: Mining and prioritization of association rules for big data: multi-criteria decision analysis approach. *J. Big Data* **4**(1), 1–21 (2017). <https://doi.org/10.1186/s40537-017-0105-4>
4. Berka, P.: Comprehensive concept description based on association rules: a meta-learning approach. *Intell. Data Anal.* **22**(2), 325–344 (2018)
5. Martin, D., Martinez-Ballesteros, M., Garcia-Gil, D., et al.: MRQAR: a generic MapReduce framework to discover Quantitative Association Rules in Big Data problems. *Knowledge-Based Systems*, 153, pp.176–192 (1 Aug 2018)
6. Rolfsnes, T., Moonen, L., Alesio, S.D., Behjati, R., Binkley, D.: Aggregating association rules to improve change recommendation. *Empir. Softw. Eng.* **23**(2), 987–1035 (2017). <https://doi.org/10.1007/s10664-017-9560-y>
7. Senthilkumar, A.: An efficient FP-Growth based association rule mining algorithm using Hadoop MapReduce. *Indian J. Sci. Technol.* **13**(34), 3561–3571 (2020)

8. Mani, K., Akila, R.: Enhancing the performance in generating association rules using singleton apriori. *Int. J. Inform. Technol. Comp. Sci.* **9**(1), 58–64 (2017)
9. Liu, Y., Xie, P., et al.: A new method based on association rules mining and geo-filter for mining spatial association knowledge. *Chin. Geogr. Sci.* **27**(3), 389–401 (2017)
10. Midaoui, O.E., Ghali, B.E., Qadi, A.E., et al.: Geographical queries reformulation using a parallel association rules generator to build spatial taxonomies. *Int. J. Electr. Comput. Eng.* **11**(3), 2586–2594 (2021)



Internet Financial News Text Classification Algorithm Based on Blockchain Technology

Long Chen(✉)

School of Economics, Shanghai University, Shanghai, China

Abstract. With the explosive growth of various Internet news text data in Internet funds, how to effectively use these text data and find out the true value behind it is very important. As the core of Internet word processing and word mining, text classification has become a major research topic in the field of natural language editing. For text classification problems, traditional methods are mainly shallow machine learning. With the rapid development of blockchain technology, huge research findings have been made in the field of image recognition and speech recognition, and the ability of blockchain to learn technical features has been further proved. This article is based on the blockchain technology model to classify Internet financial news text Research the problem. This article attempts to introduce blockchain technology related theories and design a blockchain technology model to complete the task of classifying Internet financial news texts. Overcome the shortcomings of the semantic connection between words in shallow machine learning. The final result of the research shows that the improved algorithm in this paper is 8% higher than the IG algorithm's macro average recall rate and 2% higher than the PIG algorithm. The F1 index has increased by 6% and 2% respectively. The experiment proves the feasibility and effectiveness of the improved IG feature selection algorithm proposed in this paper.

Keywords: Internet finance · Text classification · Blockchain technology · Machine learning

1 Introduction

As an important way to process documents, text classification plays an important role in information processing. In Internet finance, many news information resources are presented in the form of text. People expect to quickly and accurately obtain useful information through news text data processing. In recent years, with the gradual deepening of blockchain technology, text sorting technology has been well developed [1]. However, accurately classifying text into a large amount of text is not as simple as we think. It usually needs to go through the steps of word preprocessing, feature selection, feature weighting and sorter training. The text ranking algorithm used in these specific steps is mainly based on the research of specific feature selection, attribute weighting and classifier algorithm in Chinese text ranking, and an improved algorithm is proposed.

In recent years, many researchers have conducted research on the Internet financial news text classification algorithm of blockchain technology and have achieved good

results. For example, Wu believes that the Internet contains a wealth of information resources. It has become a worldwide library in the digital age, and it has become a means of exchanging ideas and obtaining information [2]. Liu believes that due to the openness and dynamics of the network, the distribution of information resources on the network is scattered, and there is no unified management and structure. Therefore, the utilization efficiency of information resources is obviously reduced, which leads to difficulties in obtaining information [3]. At present, domestic and foreign scholars have conducted a lot of research on the Internet financial news text classification algorithm of blockchain technology. These previous theories and experimental results provide a reliable theoretical basis for the research of this article.

Based on the relevant characteristics of blockchain technology, this article improves and analyzes the Internet financial news text classification algorithm. Due to the faster update of network data and more content, how to quickly and accurately find the required information resources from the massive data has changed A problem encountered by most users. To solve this problem requires a lot of text classification information, so automatic text classification technology has become a development trend and has become a very meaningful research direction. This article analyzes financial text classification algorithms on the basis of blockchain to provide enough Rich sample diversity, try to introduce a corpus of sufficient scale to better describe semantic features.

2 Research on the Application of Blockchain Technology in the Text Classification of Internet Financial News

2.1 Introduction to Blockchain Technology

(1) Overview of blockchain technology

Blockchain technology is essentially a distributed witness technology. The so-called distributed means that the data is not concentrated in a certain data server center, but is stored in each node in the network. The network members themselves are the storage carriers of the data, and directly share, copy, store and synchronize the data [4, 5]. The so-called “witness” refers to the confirmation and notarization of the information uploaded on this distributed network. Once the information is uploaded and verified successfully, it cannot be tampered with to achieve the purpose of “witness”. Blockchain is the storage form of blockchain technology. The blockchain is composed of “blocks” connected in chronological order, and corresponding information is recorded in each “block”. In the earliest application of blockchain technology to Bitcoin, each block records both parties of the transaction and the transaction amount, timestamp (a string of fields for recording the time), the hash value of this block and the previous block’s hash value. Hope value, etc., each block is connected with the corresponding hash value through a timestamp to form a chain [6, 7]. If the blockchain is compared to a ledger, that block is a page of the ledger, which records the corresponding transaction information.

(2) Blockchain structure analysis

The “page number information” of the block is recorded in the block chain, and its content includes: block chain version number, hash value of the previous block,

Merkle root, timestamp, difficulty target, Nonce, etc. Among them, the blockchain version number shows its applicable version number and corresponding consensus rules. The hash value of the previous block is equivalent to the number of the previous block. The Merkle root is used to quickly query the corresponding transaction, and the timestamp shows the area. The generation time of the block. The difficulty target is the number of 0s that start with a string of 0s in the hash value of the block, and Nonce is the final solution to the problem of the POW consensus mechanism of the block. How to achieve the proof-of-work (POW) consensus mechanism through the difficulty target and the Nonce value will be explained in the “consensus mechanism” below [8]. The realization of blockchain technology, combined with existing cryptographic technologies such as asymmetric encryption and hashing, has created a consensus mechanism to solve the asynchronous problem of distributed networks and constructed a brand-new witness storage technology. Among them, asymmetric encryption technology is used for user identity verification, which is equivalent to providing customers with a mechanism for registering “accounts” and “passwords”. Hash operation is used to verify transactions and realize the consensus mechanism of proof of work, and is used to ensure the consistency of information between nodes in the blockchain network.

2.2 The Consensus Mechanism of Blockchain Technology

(1) Work proof mechanism

Taking Bitcoin as an example, the Bitcoin blockchain uses a POW proof-of-work mechanism. Every time an accountant is introduced, the POW mechanism creates a block. There are multiple user nodes in the blockchain system, which are used to continuously test different Nonce values and add the Nonce value to the hash value calculation of the current block to obtain a series of hash values that meet the difficulty value, because the hash value calculation Uncertainty and irreversibility to obtain the Nonce value that satisfies the conditions requires constant trial and error calculations. These nodes that undertake the calculation work are also called miners, and the process for miners to calculate the Nonce value is mining [9, 10]. The miner who successfully solves the result is a recognized accountant, and obtains the right to the current block’s books and receives certain rewards. That is, in the next blockchain expansion, the one with the fastest expansion speed can be recognized by the entire network. This can ensure that the blockchains stored by the nodes in the entire network are consistent, and avoid paying the same bitcoin to two people at the same time. The advantage of the POW mechanism is that it is simple and direct. Attacking the blockchain consensus mechanism requires strong computing power, high cost, and strong anti-interference ability. However, the workload proof mechanism also has some problems: for example, it needs to consume a lot of energy, the computing power is gradually concentrated, and the possibility of monopolizing the bookkeeping power is high.

(2) Proof of rights mechanism

Because the Bitcoin workload proof mechanism consumes a lot of computing power and energy, and as computing power becomes more concentrated, it becomes more like a centralized system. So someone proposed a POS mechanism. The POS equity

proof mechanism, also known as the equity proof mechanism, is because under this consensus mechanism, currency holders can earn interest through currency holdings. The POS mechanism calculates the user's "coin age", that is, the number of users holding coins multiplied by their holding time. The higher the user's currency age, the higher the probability of being elected as the bookkeeper. In addition, when the accountkeeper is elected as after the bookkeeper has successfully booked the book, its currency age will be cleared to zero [11]. Such a mechanism guarantees that no node will continue to be the bookkeeper to monopolize the bookkeeping power. However, in the POS mechanism, the cost of forged blocks is low. Unlike in POW, the corresponding Nonce value must be calculated for forged blocks, which is prone to blockchain bifurcation. Ethereum is currently gradually transferring the consensus mechanism from POW to PoS.

(3) Appointment of equity certification mechanism

The DPOS appointed equity proof mechanism is similar to the board election mechanism. A certain number of nodes will be elected as nodes with the right to keep accounts in the system at regular intervals. Compared with POS and POW, DPOS's method of electing bookkeepers is more direct and concise, but some critics believe that this election method has actually returned to a centralized system, which violates the original intention of blockchain technology to decentralize [12]. The virtual currencies that use the DPOS consensus mechanism mainly include BTS and EOS. Taking EOS as an example, in the system, each token holder uses the number of virtual currency holdings as the voting weight to select 21 master nodes. These master nodes generate blocks in turn and obtain Corresponding rewards, the system will organize a re-voting every 10 rounds, and malicious or incompetent bookkeepers will be eliminated. It seems that DPOS is prone to the problem of bookkeeping monopoly, but the introduction of the elimination mechanism makes any malicious node that wants to monopolize the bookkeeping power for their own benefit will be eliminated, and the fork caused by the malicious node can also be followed by subsequent nodes. Smooth, seemingly centralized blockchain actually has higher bookkeeping efficiency.

3 Research on the Experimental Preparation for the Application of Internet Financial News Text Classification

3.1 Experimental Method

The contour index evaluation method evaluates the clustering results based on the degree of cohesion of the same attribute data and the degree of separation of different attribute data. It can be performed on different algorithms under the same basic principle of data experiment, or on the experimental results generated by different test methods under the same algorithm. Evaluation. The following is the contour factor formula.

$$SE = \frac{1}{n} \sum_{i=1}^n \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (1)$$

In formula (1), n represents the total number of data objects, $a(i)$ represents the average value of the sum of the distances between data object i and other constituent elements of this category, and $b(i)$ represents the distance between data object i and all other constituent elements of this category. And the average value. The maximum value of the contour coefficient is 1, and the minimum is -1 . The larger the value, the better the comprehensive evaluation effect of the cohesion of the same attribute data and the separation of different attribute data.

CH indicator(Calinski-Harabasz index):

$$CH(K) = \frac{Tr(S_B)}{K-1} / \frac{Tr(S_W)}{N-K} \quad (2)$$

In Eq. (2), $Tr(S_B)$ and $Tr(S_W)$ respectively represent the traces of the divergence matrix between different classes and the divergence matrix within the same class.

3.2 Experimental Data Collection

The implementation process of the experiment is mainly to investigate the application of the Internet financial news text classification algorithm of blockchain technology. The data used in the experiment includes classified treasures and news text data obtained by news portals. There is a one-to-one match between news category and URL. The Internet financial news text data is financial news text, and the classification treasure text data is officially provided by Sogou Input Method. The use of blockchain technology to analyze the classification algorithms of all categories in the educational text set, thereby verifying the effectiveness of the relevant algorithms for the classification of Internet financial news texts.

4 Analysis of Experimental Results of Internet Financial News Text Classification Application

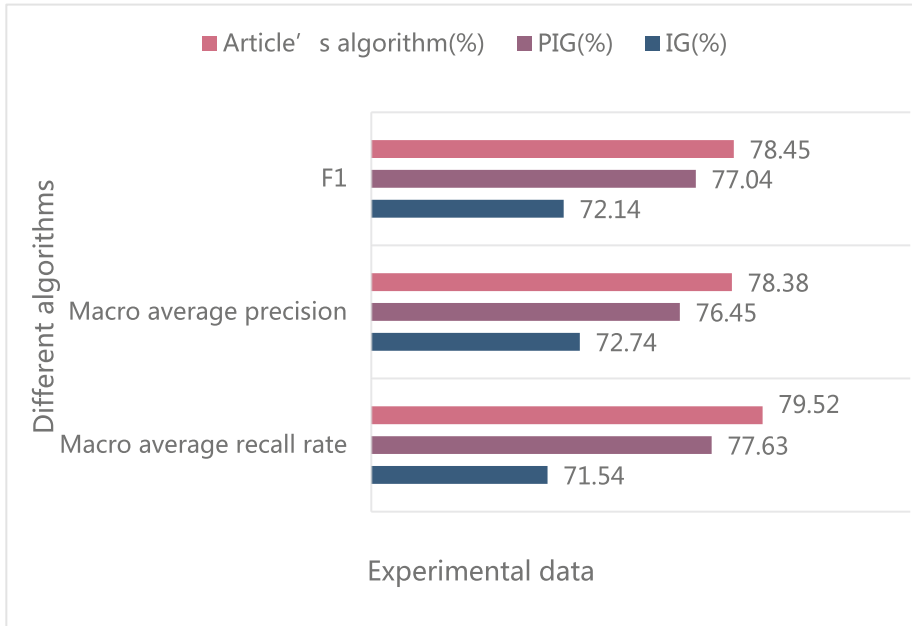
4.1 Analysis of Recalls Obtained by Different Feature Selection Algorithms

In order to test the feasibility of the feature selection algorithm for improving the information gain in this chapter, an experimental analysis is carried out. Perform 10 experiments on the data set each, and take the average of the 10 experiments as the final result. Experimental simulations have obtained the recall and precision of each category under the traditional IG algorithm, PIG algorithm and the improved algorithm in this article. The experimental results are shown in the Table 1 shown.

As shown in Fig. 1, under different feature selection algorithms, the data obtained by the improved algorithm proposed in this article is better than other algorithms. The recall and precision rates of the IG algorithm are 71.54% and 72.74%, respectively, and the data of the PIG algorithm are 77.63% and 76.45%, respectively. The gap between the IG algorithm and the PIG algorithm is not large, and the improved algorithm in this paper is 8% higher than the IG algorithm's macro average recall rate and 2% higher than the PIG algorithm. F1 the index increased by 6% and 2% respectively. The experiment proves the feasibility and effectiveness of the improved IG feature selection algorithm proposed in this paper.

Table 1. Recall analysis table obtained by different feature selection algorithms

Options	IG (%)	PIG (%)	Article's algorithm (%)
Macro average recall rate	71.54	77.63	79.52
Macro average precision	72.74	76.45	78.38
F1	72.14	77.04	78.45

**Fig. 1.** Recall analysis graph obtained by different feature selection algorithms

4.2 The Overall Evaluation and Analysis of the Naive Bayes Algorithm and the Improved Algorithm in this Paper

In order to test the feasibility of improving the naive Bayes algorithm of mutual information weighting in this chapter, an experimental analysis is carried out. Perform experimental simulation analysis on the data set, and the experimental results are shown in Fig. 2.

As shown in Fig. 2, the overall data situation of the naive Bayes algorithm and the improved algorithm in this paper can be seen, the data of naive Bayes shows that the recall rate and the precision rate are 73.74% and 78.64%, respectively, while the improved algorithm The data are 81.24% and 86.36%, which are an increase of 8% and 6% respectively compared with the naive Bayes algorithm. At the same time, the F1 index increased by about 6.5%. The experimental results prove the feasibility of the naive Bayes algorithm with weighted reciprocal information proposed in this paper,

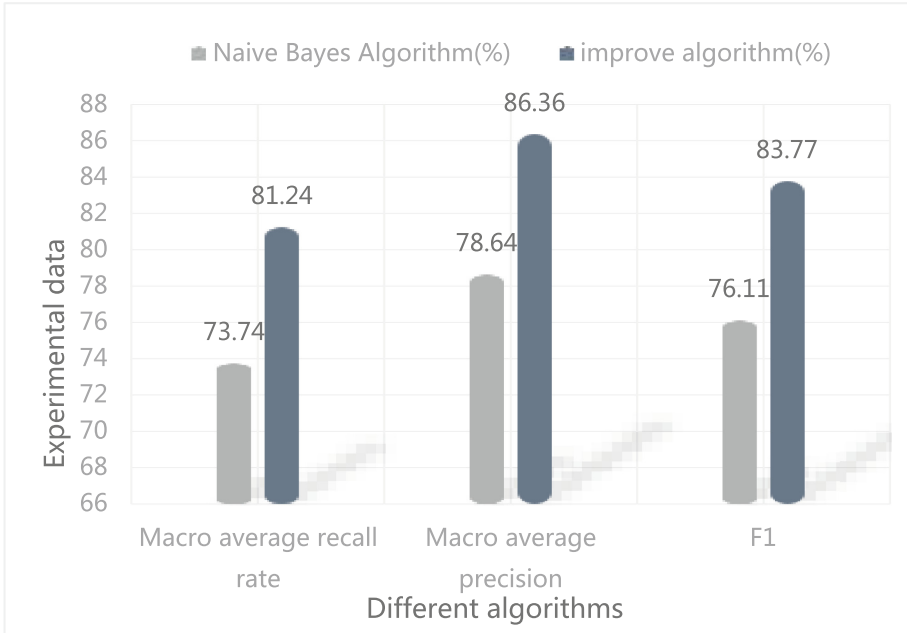


Fig. 2. Naive Bayes algorithm and the overall evaluation analysis diagram of the improved algorithm in this paper

indicating that the effect of Internet financial text data classification has been significantly improved.

5 Conclusions

The research object of this paper analyzes the research and analysis of the Internet financial news text classification algorithm under the blockchain technology, compares the traditional algorithm with the improved algorithm of this paper, and analyzes the feasibility of the algorithm through the macro average precision rate and the macro average recall rate. Today, there is an online text database containing a large amount of data on the Internet. How to perform fast and accurate text analysis on these text data is a research direction that deserves attention. Internet economic news texts focus on social hot spots and have a large number of users in China. Its potential application value is of great significance to promoting economic and social development. Today's text sorting methods are different and adapt to different data environments. Blockchain technology allows the text analysis process to adapt to complex application environments, to adapt to the variability of data and the impact of data noise, so that the data processing process remains reliable and practical. It is a very difficult question. Social media focuses on a large number of hot topics, search engines are continuously optimized according to user behaviors, the rapid response of information retrieval systems, and the high topicality of intelligence analysis are all in urgent need of analysis and development in a wider range of fields.

References

1. Liu, L., Xu, B.: Research on information security technology based on blockchain (1), 380–384 (2018)
2. Wu, X.: Research on English online education platform based on genetic algorithm and Blockchain technology. *Wirel. Commun. Mob. Comput.* **2020**(9), 1–7 (2020)
3. Liu, L., Tsai, W.T., Bhuiyan, M., et al.: Automatic Blockchain whitepapers analysis via heterogeneous graph neural network. *J Parallel Distrib Comput* **5**, 145–146 (2020)
4. Zheng, W., Wang, K., Wang, F.Y.: GAN-based key secret-sharing scheme in Blockchain. *IEEE Trans. Cybern.* **51**(1), 393–404 (2021)
5. Pandey, P., Litoriya, R.: Implementing healthcare services on a large scale: challenges and remedies based on Blockchain technology. *Health Policy and Technology* **9**(1), 69–78 (2020)
6. Jang, K.J.: The a study on innovative financial services of business models using BlockChain technology. *The e-Business Studies* **18**(6), 113–130 (2017)
7. Zhang, J.: Interaction design research based on large data rule mining and Blockchain communication technology. *Soft. Comput.* **24**(21), 16593–16604 (2020)
8. Huang, W., Dai, F.: Research on digital protection of intangible cultural heritage based on Blockchain technology. *Inform. Manage. Comput. Sci. (IMCS)* **2**, 1–13 (2019)
9. Yue, Y.: Research on credit risk early warning based on RBF neural network—taking internet financial trading platform as an example. *Model. Simul.* **10**(2), 257–267 (2021)
10. Verma, S.: Analysis on different strategies used in Blockchain technology. *J. Comput. Theor. Nanosci.* **16**(10), 1–5 (2019)
11. Wu, Y., et al.: A Bitcoin transaction network analytic method for future Blockchain forensic investigation. *IEEE Trans. Netw. Sci. Eng.* **8**(2), 1230–1241 (2020)
12. Peng, C., Bao, T.: An analysis method for interpretability of CNN text classification model. *Future Internet* **12**(12), 228–230 (2020)



Measurement and Change Analysis of Trade Implied Carbon Based on Input-Output Method and SDA Method Under Modern Information Technology

Jing Ye¹(✉) and Shalini Kumar Shukla²

¹ School of Economics, Shanghai University, Shanghai 200444, China

² KL University Vijayawada, Vijayawada, Andhra Pradesh, India

Abstract. In this document, the input-output method and the SDA method are combined within the framework of modern information technologies to measure carbon emissions and changes in trade between China and the EU. The input-output method is used to calculate the total direct and indirect carbon emissions hidden throughout the entire commodity production process. This document uses the structural decomposition method to further break down the factors influencing the growth of carbon emissions incorporated from exports to the EU in terms of energy efficiency, production technology, export scale and export structure. The contribution rate of energy use efficiency to global carbon reduction incarnate is 12.8%. Both the size and structure of the export from China to the EU promote the growth of carbon emissions embodied to varying degrees, in which the export dimension plays an important role, contributing to 104.2%.

Keywords: Sino-European Trade · Embodied carbon emission · The input-output method · Structural decomposition analysis

1 Introduction

Global warming is a major problem faced by countries around the world. According to the statement on global climate in 2020 issued by the United Nations Environment Programme in March 2021, the global average carbon dioxide concentration in 2020 is 408.04 ± 0.1 shares per million, and the average temperature is 1.1 °C higher than before industrialization. Carbon dioxide and other greenhouse gas emissions are important causes of global warming. World economic integration and the rapid growth of international trade, China, as the main exporter of manufactured products, has undertaken a considerable amount of hidden carbon emissions to other countries to clarify countries' responsibilities for energy conservation and emission reduction.

Scholars at home and abroad have conducted extensive research and Analysis on the relationship between trade and carbon emissions, mainly including the relationship between trade and carbon emissions, carbon emissions brought by trade and the influencing factors of carbon emissions reflected in trade [1–3]. Li et al. (2020) [4] pointed

out that due to Sino US trade, China's inefficient manufacturing technology and coal biased energy structure have led to an increase in global carbon dioxide emissions. Wang et al. (2019) [5] studied the carbon contained in Sino Australian trade, and the results show that the carbon contained in China's export to Australia is much higher than that imported from Australia. Iftikhar (2018) [6] analyzed the input-output model and concluded that the environmental performance of Japanese industry was much higher than that of American industry, thus reducing the global carbon emissions caused by trade between Japan and the United States.

The input-output method and the life-cycle method are mainly used to calculate the carbon content. The input-output method generally adopts the input-output model of a single region, multi-region, multi-product or sector to calculate and analyse carbon content in international trade goods [7–9]. The life cycle method mainly measures energy consumption and carbon dioxide emissions from different products at different life cycle stages [10, 11]. These calculations aim at analysing whether international trade transfers hidden carbon emissions and how much. In recent years, Some scholars began to combine the structural decomposition analysis (SDA) with the input-output model to analyze the factors of influence of carbon emissions reflected in trade.

To sum up, although many scholars have studied carbon emissions in trade from different angles and different research methods, so far, few have studied carbon emissions in China EU trade in combination with input-output model and SDA model. Therefore, Based on the existing research results at home and abroad, this paper will eliminate the impact of imported intermediate products, establish a competitive input-output model, calculate China's implied carbon emissions due to China EU trade, and make a structural decomposition analysis of four factors, trying to reveal the reasons affecting China's implied carbon emissions growth in China EU trade.

2 Estimate the Embodied Carbon Emission in the Sino-European Trade

2.1 Input-Output Model

Commodity production not only needs the energy consumption of the final production department, but also indirectly leads to the energy consumption of other departments in the production process. Therefore, when calculating the implied carbon emissions of international trade in a sector, it is necessary to calculate the direct carbon emissions and indirect carbon emissions caused by the sector at the same time. The input-output method can measure all direct and indirect implicit carbon emissions in the whole process of commodity production through the interdependence between the input and output of various departments in the econometric system.

According to the input-output method, the following mathematical relations can be obtained:

$$X + M = Y + AX + E \quad (1)$$

$$\begin{bmatrix} X^1 \\ X^2 \\ \vdots \\ X^n \end{bmatrix} + \begin{bmatrix} M^1 \\ M^2 \\ \vdots \\ M^n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n y^{1j} \\ \sum_{j=1}^n y^{2j} \\ \vdots \\ \sum_{j=1}^n y^{nj} \end{bmatrix} + \begin{bmatrix} A^{11} A^{12} \dots A^{1n} \\ A^{21} A^{22} \dots A^{2n} \\ \vdots \\ A^{n1} A^{n2} \dots A^{nn} \end{bmatrix} \dots \begin{bmatrix} X^1 \\ X^2 \\ \vdots \\ X^n \end{bmatrix} + \begin{bmatrix} E^1 \\ E^2 \\ \vdots \\ E^n \end{bmatrix} \tag{2}$$

In the Eq. (1), X is the matrix of total domestic output, M is the entry matrix, Y is the final use matrix in China, A is the technology matrix (represented by input-output table (value type)), So AX is the domestic indirect using matrix, and E is the exit matrix (Table 1).

Table 1. IRIO input-output table

Project		Intermediate demand	Final demand	Import	Export	Total output
		Each department				
Intermediate input	Each department	Z ^{RS}	Y	M	E	X
Added value		V				
The total investment		Z				

$$A = \frac{Z^{RS}}{X} \tag{3}$$

In the input-output table statistics, the raw materials of intermediate supplies not only come from domestic production, but also include imported raw materials. In order to eliminate the influence of imported intermediate products, the import coefficient matrix U is constructed. It is assumed that imported intermediate goods are interchangeable with domestic products. That is, the input-output model established in this paper belongs to the competitive type. In this paper, U is designed as a diagonal matrix. U_{ij}, as a proportional coefficient, represents the proportion of the import m_i of the i sector in the total domestic demand of the sector. The calculation formula is as follows:

$$M = U(AX + Y) \tag{4}$$

Where the elements of the diagonal matrix U:

$$u_{ii} = \frac{m_i}{\sum_{j=1}^n a_{ij}x_j + y_j} \tag{5}$$

$$\begin{bmatrix} M^1 \\ M^2 \\ \vdots \\ M^n \end{bmatrix} = \begin{bmatrix} \frac{m_1}{\sum_{j=1}^n a_{1j} + y_j} & 0 & \dots & 0 \\ 0 & \frac{m_2}{\sum_{j=1}^n a_{2j} + y_j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{m_n}{\sum_{j=1}^n a_{nj} + y_j} \end{bmatrix} \left\{ \begin{bmatrix} A^{11} & A^{12} & \dots & A^{1n} \\ A^{21} & A^{22} & \dots & A^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A^{n1} & A^{n2} & \dots & A^{nn} \end{bmatrix} \begin{bmatrix} X^1 \\ X^2 \\ \vdots \\ X^n \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^n y^{1j} \\ \sum_{j=1}^n y^{2j} \\ \vdots \\ \sum_{j=1}^n y^{nj} \end{bmatrix} \right\} \tag{6}$$

It can be obtained from Eqs. (1), (2) and (4):

$$X = D[(I - U)Y] + DE \tag{7}$$

$$D = [I - (I - U)A]^{-1} \tag{8}$$

D is Leontief inverse matrix, W is defined as the direct matrix of carbon dioxide emission coefficient.

$$w_i = \frac{\alpha_k \times q_{ik}}{x_i} \tag{9}$$

α_k represents the direct CO₂ emission coefficient of various energy sources, q_{ik}/x_i represents the energy consumption intensity of type K in the industry sector i. x_i represents the total output of product department i. q_{ik} represents the total consumption of type K energy in the product sector i.

Finally, the total implied carbon emission formula of the country is obtained:

$$C = WX = WD[(I - U)Y] + WDE \tag{10}$$

2.2 Data Description

This study concerns energy data, economic data, commercial data. α_k can be calculated in accordance with the guidance of the IPCC. This document divides energy consumption by sector into eight types. The process of calculating the carbon dioxide emission coefficient, α_k , of these eight major energy sources, is as follows:

$$\alpha_k = CEF_k \times COF_k \times NCV_k \times (44/12) \quad (k = 1, 2, 3, \dots, 8) \quad (11)$$

The α_k is the direct emission coefficient of CO₂ from different energy sources, the results of the calculation are shown in Table 2. CEF is a carbon emission factor provided by the IPCC. COF is a carbon and oxygen binding factor (Carbon oxidation factor, take the deficiency value 1). NCV means a low average calorific value of primary energy, data are available from the China Energy Statistical Yearbook. The units of crude oil, kerosene, gasoline, fuel oil, coca, diesel and coal are 10 Kt/ 10 Kt. The natural gas unit is 10 Kt/BCM (Billion Cubic Meters).

Table 2. Direct CO₂ emission coefficient of various energy sources(10 Kt/10 Kt, 10 Kt/BCM)

Types of energy	Coal	Coke	Crude oil	Gasoline	Kerosene	Diesel	Fuel oil	Natural gas
α_k	2.03	2.66	3.07	3.19	3.08	3.16	3.22	218.4

Note: The data sources in the table are calculated according to Formula (11).

The data of q_{ik} and x_i come from China Statistical Yearbook. All data in m_i and E come from China Statistical Yearbook of Foreign Trade. The data of China and EU interregional input-output tables are all from WIOD.

2.3 Conclusions and Discussion

Based on the model, we calculate the total implicit CO₂ of exports from China to the EU in 2010 to 2020. (Table 3).

Table 3. The total value of China's exports to the EU and the total CO₂ in them

Year	2010	2020
Total carbon emissions implied by exports (10 Kt)	40251.31	55086.54
Total exports (billion dollars)	3731.5	4449.7

From 2010 to 2020, there was an upward trend in total CO₂ from Chinese exports to the EU. Compared to 2010, the CO₂ produced by China's exports of raw materials to the EU increased, from 402513.1 Kt to 55865.4 Kt in 2020. In 2020, the implicit carbon

emissions of Chinese exports to the EU were 136,86% of those of 2010. China accounts for a significant proportion of carbon emissions in EU consumption.

Table 4 shows the implied carbon emissions and export amount of China's commodity sector exported to EU. The top three sectors of embodied carbon emissions from China's exports to the EU in 2010 were metal smelting and calendering industry (22.51%), general specialized equipment manufacturing (21.74%), chemical industry, plastic and rubber manufacturing (15.71%). In 2020, the top three sectors of embodied carbon emissions from China's exports to the EU were chemical industry, plastic and rubber manufacturing (24.34%), general and special equipment manufacturing (20.9%), and electrical, machinery and equipment manufacturing (9.87%).

Table 4. Implied carbon emissions from China's export goods sector to the EU

Serial number	Industry sectors	Implicit carbon emissions from exports of goods to the EU by sector (10 Kt)		Proportion of embodied carbon emissions from exports of goods to EU by sector in total emissions (%)	
		2010	2020	2010	2020
1	Agriculture, forestry, animal husbandry and fishery	195.98	210.34	0.49	0.38
2	Metal mining industry	476.55	100.34	1.18	0.18
3	Non-metallic mining and other mining industries	274.5	335.78	0.68	0.61
4	Food, beverage and tobacco processing industries	176.4	257.21	0.44	0.47
5	The textile industry	3524.18	5166.51	8.76	9.38
6	Garments, shoes, hats, leather, eiderdown and other products	433.35	891.62	1.08	1.62
7	Wood processing and furniture manufacturing	1072.35	2249.57	2.66	4.08
8	Paper printing, culture and education, sports goods manufacturing industry	1619.33	3141.95	4.02	5.70
9	Petroleum processing, coking and nuclear fuel processing industry	652.95	301.22	1.62	0.55
10	Chemical industry, plastic and rubber manufacturing	6324.08	13407.90	15.71	24.34

(continued)

Table 4. (continued)

Serial number	Industry sectors	Implicit carbon emissions from exports of goods to the EU by sector (10 Kt)		Proportion of embodied carbon emissions from exports of goods to EU by sector in total emissions (%)	
		2010	2020	2010	2020
11	Nonmetallic mineral products industry	2351.25	4567.54	5.84	8.29
12	Metal smelting and rolling processing industry	9061.65	3560.14	22.51	6.46
13	Metal products industry	1091.25	1485.88	2.71	2.70
14	General, special equipment manufacturing	8751.825	11513.32	21.74	20.90
15	Transportation equipment manufacturing	772.2	1243.59	1.92	2.26
16	Electrical, machinery and Equipment manufacturing	2478.38	5436.49	6.16	9.87
17	Instrument and cultural office machinery manufacturing industry	734.4	805.53	1.83	1.46
18	Crafts and other manufacturing	203.85	239.44	0.51	0.44
19	Others	56.86	172.17	0.14	0.31

3 Structural Decomposition Analysis of the Growth of Implied Carbon Emissions in China's Export Trade to the EU

3.1 Structural Decomposition Analysis (SDA)

The SDA is to measure the influence of each variable on the dependent variable. This paper uses the SDA to decompose the change of embodied carbon in China's export trade to EU in 2010 and 2020.

$$C = WX = WD[(I - U)Y] + WDE \quad (12)$$

$$C_E = WDE = W[I - (I - U)A]^{-1}E \quad (13)$$

$$E = QK \quad (14)$$

$$C_E = W[I - (I - U)A]^{-1}QK \quad (15)$$

There are four factors that affect the carbon emissions embodied in China’s export trade to the EU: energy efficiency, the production technology, the export scale, export structure. The Eq. (15) can be obtained by using the bipolar decomposition method (The calculation period is marked with 1 and the base period is marked with 0):

$$\Delta C_E = C_1 - C_0 = G_{\Delta W} + G_{\Delta[I-(I-U)A]^{-1}} + G_{\Delta Q} + G_{\Delta K} \tag{16}$$

$G_{\Delta W}$, $G_{\Delta[I-(I-U)A]^{-1}}$, $G_{\Delta Q}$, $G_{\Delta K}$ respectively represent the contribution value of energy utilization rate, production technology, export scale and export structure to the change of China’s embodied carbon emissions from trade to EU in 2020 compared with 2010.

$$G_{\Delta W} = \frac{1}{2}\{\Delta W[I - (I - U)A]_1^{-1}Q_1K_1 + \Delta W[I - (I - U)A]_0^{-1}Q_0K_0\} \tag{17}$$

$$G_{\Delta[I-(I-U)A]^{-1}} = \frac{1}{2}\{W_0\Delta[I - (I - U)A]^{-1}Q_1K_1 + W_1\Delta[I - (I - U)A]^{-1}Q_0K_0\} \tag{18}$$

$$G_{\Delta Q} = \frac{1}{2}\{W_0[I - (I - U)A]_0^{-1}\Delta QK_1 + W_1[I - (I - U)A]_1^{-1}\Delta QK_0\} \tag{19}$$

$$G_{\Delta K} = \frac{1}{2}\{W_0[I - (I - U)A]_0^{-1}Q_0\Delta K + W_1[I - (I - U)A]_1^{-1}Q_1\Delta K\} \tag{20}$$

3.2 Results and Discussion

Decomposition the results of implied the carbon emission growth factors in China’s export trade to EU (Table 5).

Table 5. Decomposition results of implied carbon emission growth factors in China’s export trade to EU

Influencing factor	Contribution value (10Kt)		Contribution (%)	
	The formula	The results	The formula	The results
Energy efficiency	$G_{\Delta W}$	-1415.39	$\frac{G_{\Delta W}}{\Delta C_E}$	-12.8
Production technology	$G_{\Delta[I-(I-U)A]^{-1}}$	-993.96	$\frac{G_{\Delta[I-(I-U)A]^{-1}}}{\Delta C_E}$	-6.7
Exports to the EU	$G_{\Delta Q}$	15458.31	$\frac{G_{\Delta Q}}{\Delta C_E}$	104.2
Export structure to EU	$G_{\Delta K}$	786.27	$\frac{G_{\Delta K}}{\Delta C_E}$	5.3

As can be seen in Table 5, among the four decomposition factors that affect the growth of the carbon emissions that embodied China’s export trade to the EU, improvement of energy utilization efficiency and improvement of production technology contribute to concrete carbon emission reduction.

4 Conclusion and Countermeasures

According to the empirical results of input-output model and structural decomposition analysis, the following conclusions and suggestions can be drawn:

4.1 Conclusion

- (1) The carbon emissions embodied by China's exports to the EU are on the rise. China is responsible for the bulk of the carbon emissions implied by EU consumption.
- (2) Mechanical and electrical products are the main export commodities of China to EU, and also the main commodity of hidden carbon emission of China's export to EU.
- (3) Improvements in energy efficiency and production technology in China contributed to the reduction of CO₂ from exports to the EU. Export scale and export structure result in increased carbon emissions by exporting goods.

4.2 Policy Suggestions

- (1) EU and other developed consumer countries should take the initiative to undertake carbon emission reduction responsibilities and obligations.
- (2) China should further optimize the mix of exports and take full use of energy policy guidance to guide export firms and to guide exporting firms to produce more products with low energy consumption and high added value.
- (3) China should improve relevant laws and regulations to promote the introduction of advanced energy-saving and emission reduction technologies from abroad.

References

1. Rhee, H., Chung, H.: Change in CO₂ emission and its transmissions between Korea and Japan using international input-output analysis. *Ecol. Econ.* **58**, 788–800 (2016)
2. Hayami, H., Nakamura, M.: CO₂ Emission of all Alternative Technology and Bilateral Trade between Japan and Canda: Relocating Production and an Implication for Joint Implementation. Keio University, Tokyo (2012)
3. Peters, G.P., Hertwich, E.G.: CO₂ Embodied in international trade with implications for global climate policy. *Environ. Sci. Technol.* **42**, 1401–1407 (2008)
4. Zhang, H.R., Zhang, Z.K., Wei, W.D.: Regional determinants of China's consumption-based emissions in the economic transition. *Environ. Res. Lett.* **15**(7), 74–100 (2020)
5. Li, Y.L., Chen, B., Chen, G.Q.: Carbon network embodied in international trade: global structural evolution and its policy implications. *Energy Policy* **139**(4), 111–316 (2020)
6. Wang, S., Zhao, Y., Wiedmann, T.: Carbon emissions embodied in China-Australia trade: a scenario analysis based on input-output analysis and panel regression models. *J. Clean. Prod.* **220**(5), 721–731 (2019)
7. Iftikhar, Y., Wang, Z., Zhang, B.: Energy and CO₂ emissions efficiency of major economies: a network DEA approach. *Energy* **147**(3), 197–207 (2018)
8. Herrala, R., Goel, R.K.: Global CO₂ efficiency: country-wise estimates using a stochastic cost frontier. *Energy Policy* **45**(6), 762–770 (2012)

9. Wang, Q., Han, X.Y.: Is decoupling embodied carbon emissions from economic output in Sino-US trade possible? *Technol. Forecast. Soc. Chang.* **169**, 120–150 (2021)
10. Yan, Y., Yang, L.: China's foreign trade and climate change: a case study of CO₂ emissions. *Energy Policy* **38**, 350–356 (2010)
11. Gallego, B., Lenzen, A.: Consistent input-output formulation of Shared consumer and producer responsibility. *Econ. Syst. Res.* **17**, 365–391 (2005)



The Application of Big Data in Network Security and Intelligence Analysis

Wang Zheng^(✉) and Juan Liu

College of General Education,
Sichuan University of Culture and Arts, Mianyang, Sichuan, China

Abstract. With the rapid development of Internet technology, network security (NS) issues have attracted people's attention, and the types of threats to network security have become more diverse. Therefore, the application of new technologies is essential to accurately understand, monitor and analyze the network information security status in real time. However, for data sets with huge information and complex changing trends, it is impossible to completely rely on traditional manual analysis methods to reveal the effective intelligence hidden in massive data and information. This requires the use of special analysis tools, namely big data technology, to mine potential intelligence information. This article uses big data technology to construct an intelligence analysis system, organizes the data information of the perpetrators according to the police intelligence analysis work, and clusters the personnel with similar characteristics of the crime. According to the results of the cluster analysis, the perpetrators are grouped into 4 categories. Category 0 accounted for 25%, category 1 accounted for 46.67%, category 2 accounted for 18.33%, and category 3 accounted for 10%.

Keywords: Network security · Intelligence analysis · Big data technology · Cluster analysis

1 Introduction

With the continuous acceleration of the process of networking and informatization, the Internet is becoming more and more inseparable from political, economic, and cultural life, and various Internet applications are integrated into all aspects of people's lives. But at the same time, due to the complex structure of the Internet, even if the various security systems used on the network have the ability to analyze network security data, it is difficult for the network security system to transmit and store data with scattered information and large format differences. Therefore, the use of big data (BD) technology to process massive data functions to strengthen network security, build a harmonious network environment, and also provide technical support for intelligence analysis (IA).

Many scholars have achieved good results in the exploration of BD applications in NS and IA. For example, some Western countries have implemented surveillance on mobile networks. Although these surveillances did not truly monitor the security of the public Internet, and did not provide users with public Internet surveillance services, they

noticed the gaps in Internet security surveillance and the existing market. Centralized security management (SM) and control technology, there have been many concepts that regard security monitoring and security hosting as business models. Wireless operators in Western countries have begun to establish SM and control platforms around the world, ready to provide SM and control services to customers around the world [1]. In view of the large amount of data stored in the database and the increasing complexity of user query requirements, many software manufacturers have begun to develop their front-end products to solve the shortcomings of the related database decentralized public applications, through a special data synthesis engine, plus more direct data access interface, try to standardize the decentralized public application logic, quickly respond to the user's query needs, and solve the problem of data query for users [2]. Although the research results on the application of BD in NS and IA are good, in order to ensure the quality of NS and promote the efficiency of IA, advanced technology must be introduced into the actual application of NS monitoring and national IA go.

This article describes the threat factors of NS and the related functions of the IA system. The intelligence analysis structure system is designed using big data technology, and the system is applied to the public security IA. The k-means algorithm is used to deal with the complex data of the perpetrators. The information is clustered to help IA organize the information of the perpetrators and quickly analyze the characteristics of the perpetrators.

2 Network Security and Intelligence Analysis Based on Big Data

2.1 Analysis of Cyber Security Threats

(1) Threats to network security

With the development of the Internet, threats to NS have become increasingly diverse. Recently, complex and multi-directional APT attacks have been launched against specific targets, which can last for days, weeks, months or even longer. The APT attack initially collects data and intelligence information, including the collection of technical information and personal intelligence information, to pave the way for the final later attack. Since then, through various integrated attack methods, including a large number of social engineering methods, to achieve the final penetration of the target [3].

(2) Network Trojan

Network Trojans use special programs to control other computers, usually with two execution programs, the control end and the controlled end. The computer running the Trojan horse program has one or more open ports. Intruders can use these open ports to invade the computer system, and the personal information stored in the system is exposed. In order to prevent the detection of Trojan horse programs, Trojan horse developers usually use various methods to disguise the Trojan horse [4, 5].

(3) Botnet

The definition of botnet is used to help people understand the characteristics of this destructive network. Attackers use various methods to distribute bots in a distributed manner, so that most hosts on the Internet are infected, and the infected hosts receive the attacker's commands from the control channel to form a botnet [6].

2.2 Functions of Intelligence Analysis System

- (1) OLTP transaction processing NS intelligence business information and file information, and the data of the place where the data is generated is directly input into the computer. The work is mainly compartmentalized three parts: one part is the data generated in the daily work of receiving the police, filing a case, investigating, and transferring according to the intelligence business process, the other part is the use of various legal documents, and the third part is mainly the electronic equipment management [7, 8].
- (2) In view of the separation characteristics of data collection, according to the design requirements of the data warehouse and the data collection system used by the embedded data collection algorithm, a separate data collection module is set up for data preprocessing, and the quantitative data in the OLTP database (such as Case personnel, crime tools and information), qualitative (such as the nature of the case, inspection reports, etc.), and descriptive (such as site survey reports) work information through data extraction and data conversion, reorganized and stored in the DM [9].
- (3) Data Mart the purpose of establishing a data mart is to provide a comprehensive, detailed and unified data source for OLAP and DM without affecting OLTP transaction processing. Because many public security work information records have strong continuous value, we divide the logical structure of the data mart into three layers, namely, the basic data, the comprehensive data, and the thematic data layer. And establish a metadata database, describe the OLTP source database and metadata environment, formulate data processing rules, and combine advanced technology in data description. The metadata set mainly uses intelligence as a unit to establish several types of intelligence such as criminal, public security, and economics, and different types of sub-information are established in intelligence information [10].
- (4) The model library of the intelligence information analysis system is used to store data mining methods and show the relationship between models and algorithms; the specific algorithms of various methods are stored in the method library in the form of plug-ins; data mining is carried out under the guidance of the knowledge base, the obtained knowledge is stored in the knowledge base after expert inspection and analysis, for intelligence personnel to analyze or transmit to the DM database [11].
- (5) Data display presents all types of information in the form required by intelligence analysts. For example, the change curve of specific statistical indicators of specific types of cases in a specific time period, the statistical coefficients of specific quantitative and qualitative indicators, etc. Based on various data analysis and case references, intelligence analysts can propose corresponding solutions for some difficult scenarios, and even discover some undiscovered laws, which achieves the purpose of computer-aided intelligence analysis [12].

2.3 K-Means Algorithm in Intelligence Analysis

First, randomly select k cluster objects, and treat these objects as the initial cluster centers (CC). As for the remaining cluster objects, according to their distance from each CC,

they are assigned to the cluster with the closest feature, and then the cluster position of each new cluster is recalculated. Repeat this process until the standard is astrigent.

$$E = \sum_{i=1}^k \sum_{p \in c_i} |p - m_i|^2 \quad (1)$$

Among them, k is the number of clusters, E is the sum of squared errors of each object, p is a point, i represents a given object, and m_i is the mean value of cluster c_i (p and m_i are multi-dimensional).

Some data that can be represented by numerical values need to be normalized, such as the age-specific data of criminals; the normalization can use the minimum and maximum normalization methods to linearly change the original data. It is assumed that $\min A$ and $\max A$ are the minimum and maximum values of attribute A , respectively, and V is the linear conversion rate. The formula is as follows:

$$v = \frac{v - \min A}{\max A - \min A} (new_ \max_A - new_ \min_A) + new_ \min_A \quad (2)$$

3 Research on the Application of BD in NS and IA

3.1 Purpose of Inquiry

Information technology is widely used in intelligence work, which enables intelligence workers to collect a lot of social information data in actual work. These data messages contain information that we have not yet discovered. This article takes public security IA as an example, uses modern technology to process public security intelligence information, and provides important decision-making references for effectively combating cyber invaders and analyzing intelligence data, so that public security intelligence analysis work conforms to the development direction of today's digital intelligence. Use the role of big data technology to improve management analysis work of intelligence analysts.

3.2 Inquiry Method

Based on the BD technology, this paper establishes the structure model of the IA system, and uses data analysis tools to perform k-means clustering on the crime information of public security intelligence to help intelligence analysts quickly process the crime information.

4 Based on the Application of BD in IA

4.1 Structure Design of Intelligence Analysis System

Figure 1 shows the structure model of the intelligence analysis system designed using big data technology. The OLTP online analysis page interacts with data analysis and

data management systems to process and share multi-dimensional information. Multi-dimensional data analysis takes into account people’s ways of thinking, which is beneficial for intelligence analysts to interactively analyze data reflecting social security characteristics from different angles, discover the value of hidden information, avoid people from making comprehensive errors when viewing data and information, and reflect public The true face of security supports managers in decision-making on distributed heterogeneous data from information sources. The system uses OLAP tools and DM tools for data collection, stores the collected data in a database, and can extract information from the database at any time to obtain effective data.

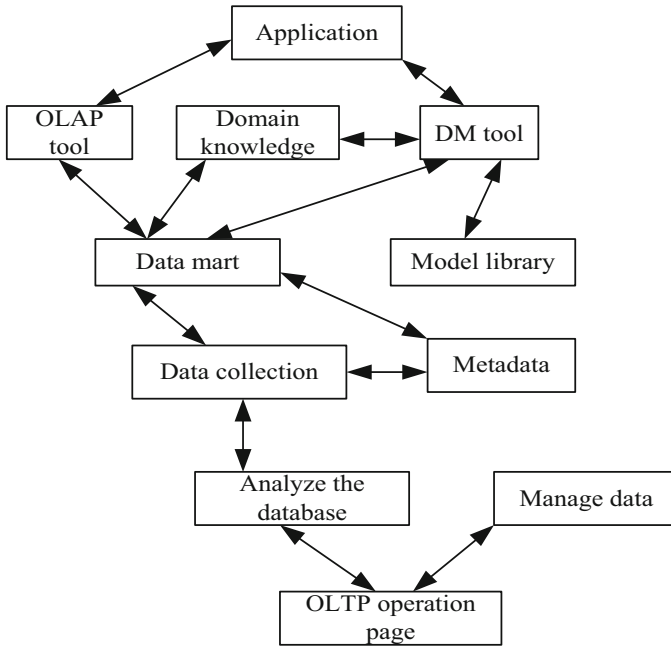


Fig. 1. Intelligence analysis system structure

4.2 Analysis of Intelligence Clustering Results

This article applies the established intelligence analysis system structure model to the public security intelligence analysis work, and analyzes the 60 criminal data information of the public security intelligence, according to the age, gender, family conditions, education level, occupation and presence of the case personnel criminal records and other characteristics are categorized, and similar characteristics will be grouped into one category. As shown in Table 1 and Fig. 2, the data sets of these 60 criminals are clustered. From the table, we can see that they are clustered into 4 categories. Category 0 means that the characteristics are those over 40 years old, the gender is girls, poor family conditions, they have a primary school education (PSE) and have no formal occupation,

5 people are classified as category 0, accounting for 25% of the total. Type 1 refers to the criminal characteristics (CC) between the ages of 30 and 40. They are all women, and the family conditions are okay. The number of people who gather in the first category is 28, accounting for 46.67% of the total. The second category refers to the CC between the ages of 18 and 40. the offender has a PSE. There are 11 people in the cluster 2 category, accounting for 18.33% of the total number of people. The CC represented by the three categories are those between 30–40 years old, Their family conditions are moderate, and they basically have high school education. There are 6 people in the three categories, accounting for 10% of the total number. It can be seen that the clustering effect of the IA system is better, and it can group crime information with similar characteristics into one category, and improve the efficiency of intelligence personnel’s analysis work.

Table 1. Display table of clustering results of similar persons

Category	Number of cases	Percentage
0	15	25%
1	28	46.67%
2	11	18.33%
3	6	10%

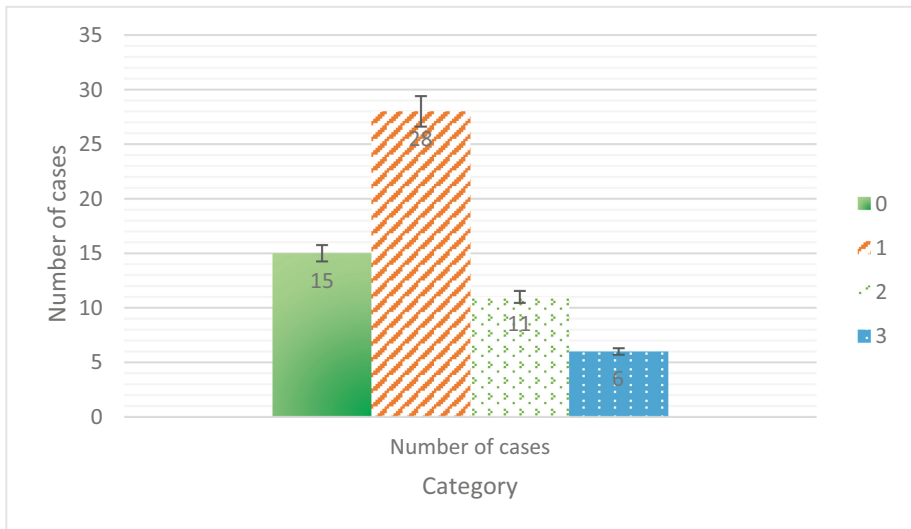


Fig. 2. Statistics of clustering results

5 Conclusion

In the era of BD, the network environment has undergone qualitative changes. NS research pays more attention to the status of the entire NS and future trends. The application research of ia based on big data in this paper shows that the use of big data technology to extract and understand complex data can help intelligence personnel to sort out the data intersection of complex crime information and improve the efficiency of dealing with network security related issues.

References

1. Kou, G., Wang, S., Tang, G.: Research on key technologies of network security situational awareness for attack tracking prediction. *Chin. J. Electron.* **28**(001), 162–171 (2019)
2. Kou, G., Wang, S., Tang, G.: Research on key technologies of network security situational awareness for attack tracking prediction. *Chin. J. Electron.* **28**(01), 166–175 (2019)
3. Roobini, M.S., Kedar, T.K., Sivasangari, A., et al.: Self-intelligence with human activities recognition based in convolutional neural network. *J. Comput. Theor. Nanosci.* **17**(8), 3484–3490 (2020)
4. Yang, H., Zeng, R., Wang, F., et al.: An unsupervised learning-based network threat situation assessment model for internet of things. *Secur. Commun. Netw.* **2020**(9), 1–11 (2020)
5. Raknys, A.V., Gudelis, D., Guogis, A.: The analysis of opportunities of the application of big data and artificial intelligence technologies in public governance and social policy. *Socialinė Teorija Empirija Politika ir Praktika* **22**(6), 88–100 (2021)
6. Zhang, Z., Song, X., Liu, L., et al.: Recent advances in blockchain and artificial intelligence integration: feasibility analysis, research issues, applications, challenges, and future work. *Secur. Commun. Netw.* **2021**(3), 1–15 (2021)
7. Mathew, A.: Artificial intelligence and cognitive computing for 6G communications & networks. *Int. J. Network Security* **10**(3), 26–31 (2021)
8. Wei, G., Sheng, Z.: Image quality assessment for intelligent emergency application based on deep neural network. *J. Vis. Commun. Image Represent.* **63**, 102581.1–102581.9 (2019)
9. Chen, Z.: Research on accounting intelligence system modeling of financial performance evaluation. *Secur. Commun. Netw.* **2021**(1), 1–9 (2021)
10. Ncubukezi, T.: A proposed: integration of the Monte Carlo model and the Bayes network to propose cyber security risk assessment tool for small and medium enterprises in South Africa. *Int. J. Comput. Sci. Inform. Secur.* **3**(18), 152–155 (2020)
11. Zhang, H., Yi, Y., Wang, J., et al.: Network security situation awareness framework based on threat intelligence. *Comput. Mater. Continua* **56**(3), 381–399 (2018)
12. Kshirsgar, P.: Deplete and discernment of security hazards in Mobile Ad-Hoc-Networks (Manet) using artificial intelligence. *Int. J. Future Gener. Commun. Netw.* **14**(1), 925–942 (2021)



Macroeconomic Forecast Based on Neural Network Algorithm and Genetic Algorithm

Yuanjie Wang^(✉)

Shanghai University, Shanghai, China

Abstract. When macroeconomic policy makers make decision-making goals, maintaining the same price level and maintaining economic growth are the best combination of goals. Throughout the history of economic growth in various countries, we can see that economic growth has always revolved around equilibrium. Therefore, policy makers need to accurately predict and ensure the changing trends of future macroeconomic dynamics and formulate prudent monetary and fiscal policies. This article focuses on the research of macroeconomic forecasting based on neural network algorithm and genetic algorithm. Based on the literature, it understands the related theories of macroeconomic forecasting and neural network algorithm and genetic algorithm, and then discusses the macroeconomic forecast based on neural network algorithm and genetic algorithm. The economic forecasting model is constructed, and the constructed model is tested. The test results show that the forecast error of the model constructed in this paper is within 10%, which shows the feasibility of the model constructed in this paper.

Keywords: Neural network · Genetic algorithm · Macroeconomics · Forecasting model

1 Introduction

After the reform and opening up, China's market economy grew out of nothing, and continued to grow rapidly. Today, it is now above the world's average value curve and has become the world's second largest economy [1, 2]. However, in the process of economic integration into economic globalization, the factors that affect economic dynamics are also affected, and it becomes more and more complex and uncertain [3, 4]. This not only caused some blindness in the formulation of national macroeconomic policies, but also brought huge pressure and rare opportunities to enterprises. In this context, macroeconomic policymakers hope to formulate scientific and wise macroeconomic control policies to promote and deepen the reform of China's socialist market economic system [5, 6]. Business decision-makers also want to accurately understand market economic trends and provide support to companies, so effective forecasting of macroeconomic dynamics is particularly important for coping with growth opportunities [7, 8].

In view of the macroeconomic forecasting research, some researchers have proposed a method of establishing a dynamic binary sequence model, extending the standard probability model to various forms of dynamic probability model, and copying the dynamic

probability model, and applying this method to the study of U.S. Treasury bonds Yield curve, the study found that this method has the ability to predict the macroeconomic [9]. Some researchers believe that the financial structure itself is very unstable and will inevitably lead to periodic financial crises, and put forward the “financial instability hypothesis” based on this. In this theory, financial structure and business cycle are inter-related. When the economy is booming, instability increases due to a series of phenomena such as currency surplus, credit expansion, and rising asset prices. With the continuous turbulence of the financial and economic structure, the macroeconomic situation will appear unbalanced, leading to further recession and even economic crisis [10]. Some scholars subdivide China’s financial market into stock markets, interbank lending markets, bond markets, and foreign exchange markets, and use out-of-sample forecasting techniques to predict future volatility or compound volatility in these markets to verify whether the method can predict economic trends. Empirical research shows that stock market volatility can be used to predict the level of economic growth and price volatility in two years. The volatility of the foreign exchange market can explain one year and the inter-bank lending market affects China’s macro economy [10]. To sum up, due to the importance of macroeconomic volatility and bond market volatility, macroeconomic forecasts have attracted much attention in academia, but there are still some problems in macroeconomic forecasts, such as the reasonable choice of forecast indicators and forecast accuracy issues are worthy of further investigation.

This paper studies the macroeconomic forecast based on neural network algorithm and genetic algorithm, analyzes the predictability of macroeconomics and the application advantages of neural network algorithm and genetic algorithm based on literature data, and then analyzes the application advantages of neural network algorithm and genetic algorithm. The macroeconomic forecasting model of the algorithm is constructed, and then the model constructed by the experiment is tested, and the relevant conclusions are drawn through the test results.

2 Macroeconomic Forecasting Neural Network Algorithm and Genetic Algorithm Research

2.1 The Predictability of the Macro Economy

In the final analysis, the predictability of macroeconomics is the predictability of human economic behavior and other natural constraints based on natural laws. People’s financial behavior can be roughly divided into two categories: individual decision-making behavior and multiplayer game behavior. Among them, individual decision-making behavior includes consumer behavior and production behavior. The hypothesis based on rational economic man can predict what people call “equilibrium” economic behavior. E.g:

- (1) Combined consumption (demand function) that maximizes consumer utility.
- (2) The balanced combination of input factors (conditional factor demand function) that maximizes the output of production factors.
- (3) Equilibrium (supply function) production that maximizes the profit of the producer.
- (4) Supply and demand balance (Walrasian general equilibrium price-general equilibrium and non-Walrasian prices).

- (5) Nash balance and improvement of micro and Hongbo, non-Nash balance (strategic balance, cooperation balance, etc.).

2.2 The Significance of Macroeconomic Forecasts

- (1) Macroeconomic forecasting is an important tool for national and regional research and formulation of plans and plans. It can not only review the strengths and weaknesses in the past economic construction, but more importantly, provide a reference for formulating future economic development strategies. In policy simulation, economic structure Analysis, corporate development strategy research and many other aspects play an important role.
- (2) Through the study of macroeconomic forecasting, the standardization and accuracy of local macroeconomic and financial management can be improved, and for government functional departments to forecast macroeconomics, control financial risks, provide management and decision-making tools, and promote the rapid development of regional e-government affairs develop.
- (3) Macroeconomic forecasts are an important basis for the government to formulate guidelines and policies, prepare and inspect plans, and adjust economic structure. Correctly use macroeconomic forecasts to strengthen the government to effectively transform the government's economic functions, strengthen the government's social management and public service functions, regulate the government's intervention in the market and intervene in economic activities, and bring the government's management of the economy into a scientific track.

2.3 Application Advantages of Neural Network Algorithm and Genetic Algorithm

- (1) The artificial neural network modeling method is different from the traditional method. Because the traditional quantitative forecasting method is a forecasting method that completely relies on big data technology, it needs complete and clear original data; the real statistical data is usually not complete and vague. However, the neural network can use its own structural adaptation process to the complex data environment (such as incomplete and inaccurate information sources) to obtain a large amount of characteristic data, so as to make a reasonable estimate of the future situation.
- (2) The artificial neural network is simple in construction, the physical meaning is clear during the training process, and the sensitivity is very strong, but the most important thing is that it can better achieve the relevant conditions of function approximation, so the imaging power is strong.
- (3) Once the traditional forecasting model is formed, its structure has great stability problems, but the realistic forecasting environment is often complicated, and the traditional forecasting model cannot satisfy the complex and changeable forecasting environment. The neural network model is a variable structure model with a very strong learning function. It can not only learn new knowledge samples through the network, but also adapt to the structure of the network, and then adapt to the changes of social economic variables.

2.4 Neural Network Algorithm and Genetic Algorithm

The overall distribution of neural network connection weights includes all the knowledge of the neural network system. The traditional weight acquisition method is to adopt a certain weight change rule, gradually adjust during training, and finally obtain a better weight distribution. For the training of feed forward neural networks, the BP algorithm is one of the most commonly used training methods. However, because the BP algorithm is a gradient descent search method, it inevitably has inherent shortcomings, such as easy to fall into the local error function extremum points, and for larger search space, multi-peak and non-biting differential functions can not effectively search for the global minimum point, and GA is an effective solution to overcome this problem, mainly because GA is a global optimize the search algorithm, so that it can avoid local minima, and there is no need to provide the gradient information of the problem to be solved in the evolution process. Training recurrent neural networks and networks with non-differentiable transfer functions have great potential.

The genetic algorithm is based on Darwin's "survival of the fittest" theory. It receives high-quality genes into the next gene through replication, mating, mutation and other characteristics, and guides it through a suitable fitness function. From several generations, we finally cultivate high-quality people, and finally cultivate high-quality people who have a higher chance of survival. Obviously, this method can be used to select initial variables. Based on this idea, some researchers have proposed a combination model of genetic algorithm and artificial neural network.

In their model, the fitness function is the reciprocal of the sum of squared errors between the output value of the neural network and the actual value.

$$f(x) = 1/mse \quad (1)$$

$$mse = \sum_{i=1}^n (p_i - t_i)^2 \quad (2)$$

In the formula, x represents the chromosome, p is the actual value of the variable target, t is the predicted value, and mse is the sum of squared errors. Under the supervision of this fitness function, it is obviously more likely to choose a combination of variables with smaller prediction errors, but variables that have little relationship with the prediction results will be eliminated.

3 Macroeconomic Forecast Based on Neural Network Algorithm and Genetic Algorithm

3.1 Selection of Predictors

(1) Inflation rate

In general, economic growth and inflation are generally moving in the same direction. Economic growth is fast, inflationary pressure is rising, and inflation is rising too fast in the process of economic growth. It is often that fixed assets have a greater

impact because it reduces inflation and inevitably affects economic growth. Different types of investors, such as credit bureaus, can pay attention to different price indices, conduct investigations through inflation control, and determine investment timing and strategies.

(2) GDP index

Due to some reasons for the Chinese economy, the national economic accounting divides GDP into three parts: consumption, investment and net exports. Government consumption is not mentioned separately, and government consumption is included in consumption and investment. Analyzing the components of GDP can roughly judge the trend of macroeconomic management, and the total retail sales of social consumption can roughly judge the upward trend of final consumption throughout the year. The total investment in fixed assets is the total annual investment.

(3) Business cycle index

The economic cycle is the process of alternate expansion and contraction of the entire economic activity. The main indicators of the economic cycle are GDP and its growth rate. One is the alternating process of rising and falling GDP or the absolute value of total production, and the other is the alternating process of rising and falling, or slowing down the relative speed of economic growth. China is a developing country, and it is unlikely that the GDP will fall absolutely. Therefore, the basis for studying China’s economic cycle is the fluctuation of GDP growth.

3.2 GA-BP Neural Network Combination Model.

(1) Network structure

The network structure diagram adopted is shown in Fig. 1:

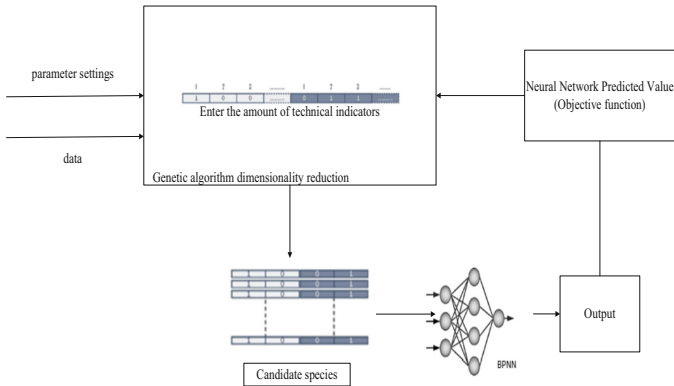


Fig. 1. Network structure

(2) Selection of the number of hidden layers of BP neural network

For a multi-level neural network, the first decision is to choose a hidden layer. Some researchers use hidden grids to approximate the continuous function in any closed space when each node has a different threshold, thus proving that the three-level neural network can go from any N dimension to another dimension, and this structure can complete the mapping M. Moreover, increasing the number of hidden layers can improve the processing capacity of the artificial neural network, but it inevitably complicates the training, increases the number of training samples, and increases the training time. Generally speaking, first define a hidden layer and add hidden layers as needed. Through logical structure and appropriate weights, the 3-level neural network can approximate complex continuous functions arbitrarily. Therefore, in this article, the GDP prediction model uses a three-layer BP neural network with only one hidden layer.

(3) Determination of the number of neurons in the hidden layer

Choosing the number of hidden layer nodes is very important in the artificial neural network model and has a significant impact on the performance of the standard neural network model. Too few hidden layer nodes will result in no network training or poor network performance. At present, there is no scientific or universal decision-making method in theory. In order to ensure a fairly high network efficiency and versatility, the most basic principle for determining the number of hidden layer nodes is to meet the accuracy requirements, that is, as few as possible, that is, to use a compact structure. If you want to specifically identify hidden-level drives, the effective method is to install a sufficient number of hidden drives first, and gradually remove non-functional hidden drives after learning, or if the network performance is poor, follow the opposite steps. In other words, the reasonable number of hidden layer nodes should be determined by the way the nodes are removed and expanded, while considering the complexity of the network structure and the size of the error. Through experiments, the hidden layer of the network model established in this paper is more suitable to get 6 nodes.

3.3 Establish a GA-BP Neural Network Combined Model for China's Macroeconomic Forecasting

The GDP forecast model established in this paper has three levels: input level, hidden level and output level.

- (1) Input: As mentioned in the above question, the n-th year economic growth rate, inflation rate and business cycle rate are used as input level information.
- (2) Hidden layer: Experience has shown that the hidden layer selects 6 neurons.
- (3) Output: The neural network model has only one output, that is, $n + 1$ year GDP.
- (4) Excitation function: adopt sigmoid type excitation function.

The purpose of introducing genetic algorithm is to try to optimize the input variables of neural network.

3.4 Model Evaluation

Judge the production efficiency of the combined model calculation, mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), mean absolute error rate (MARE), mean square percentage error (MSRE) for comparison and testing ten statistical indicators (also called loss functions).

4 Model Detection

4.1 Source of Data

Regarding data selection. Through the above detailed and clear introduction and explanation of the characteristics and advantages of the artificial neural network model, when doing macroeconomic modeling, it is necessary to train the neural network with historical data, so that a more accurate financial forecast model can be trained (this article In, the data comes from the economic value of World Net’s GDP from 2015 to 2019).

4.2 Selection of Parameters

Determining the parameters of a genetic algorithm requires a lot of experimentation to finally determine relatively reasonable values. This genetic algorithm system adopts {0,1} binary coding method, code length L is 45, initial population size N is 21, maximum genetic algebra T is 55, roulette selection probability S is 0.8, crossover probability C is 0.8, mutation probability M is 0.01. BP neural network parameters are selected as above.

Since the error of training samples can be small, test samples are randomly selected from aggregated samples to represent the accuracy of network calculation and prediction (network performance). The prediction results are shown in Table 1.

Table 1. Model prediction results

	Actual value	Predictive value
2015	89560.9	88980.5
2016	98782.9	95974.7
2017	109797.7	108655.8
2018	123430.6	121749.7
2019	145494.4	148601.8

As can be seen from Fig. 2, because the error between the predicted result and the actual result is within 10%, the neural network model based on artificial neural network and genetic optimization algorithm is more accurate in predicting GDP. This shows that economic growth rate, currency expansion rate and business cycle are feasible as forecasting models for establishing neural network models.

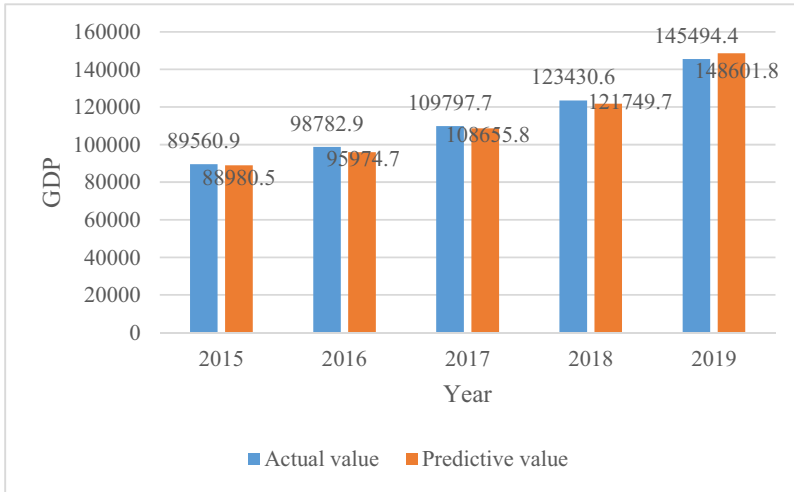


Fig. 2. Model prediction results

5 Conclusions

This paper studies the macroeconomic forecasting based on neural network algorithm and genetic algorithm. After understanding the relevant theories, constructs the macroeconomic forecasting model based on neural network algorithm and genetic algorithm, and validates the constructed model. As a result, the prediction results of the model are more accurate, so the prediction mode of the model is feasible.

References

1. Hou, C.: Infinite hidden markov switching VARs with application to macroeconomic forecast - ScienceDirect. *Int. J. Forecast.* **33**(4), 1025–1043 (2017)
2. Bd, A., Ci, B., Kk, C.: High-frequency credit spread information and macroeconomic forecast revision - ScienceDirect. *Int. J. Forecast.* **36**(2), 358–372 (2020)
3. Jo, S., Sekkel, R.: Macroeconomic uncertainty through the lens of professional forecasters. *J. Bus. Econ. Stat.* **37**(3), 436–446 (2019)
4. João, V., Jalles, J.T.: Model-based vs. professional forecasts: implications for models with nominal rigidities. *Macroecon. Dyn.* **21**(01), 130–159 (2016)
5. Pedersen, M.: Anomalies in macroeconomic prediction errors—evidence from Chilean private forecasters. *Int. J. Forecast.* **35**(3), 1100–1107 (2019)
6. Shirov, A.A., Gusev, M.S., Frolov, I.E.: Macroeconomic effects of russian defense expenditures: retrospective analysis and forecast. *Stud. Russ. Econ. Dev.* **29**(4), 343–351 (2018). <https://doi.org/10.1134/S1075700718040159>
7. Ilyukhin, A.A., Ponomaryova, S.I., Ilyukhina, S.V.: Macroeconomic analysis of Sverdlovsk oblast budget forecast in the context of economic growth in Russia. *Upravlenets* **70**(6), 72–79 (2017)
8. Baetje, F., Friedrici, K.: Does cross-sectional forecast dispersion proxy for macroeconomic uncertainty new empirical evidence. *Econ. Lett.* **143**, 38–43 (2016)

9. Kapetanios, G., Hyndman, R.J.: A comprehensive evaluation of macroeconomic forecasting methods. *Int. J. Forecast.* **35**(4), 1226–1239 (2019)
10. Broughton, J.B., Lobo, B.J.: Herding and anchoring in macroeconomic forecasts: the case of the PMI. *Empir. Econ.* **55**(3), 1337–1355 (2017)



Intelligent Optical Distribution Technology in Power Communication Transmission

Xu Zhang^(✉), Xin Xiao, Le Ma, Chao Liu, Xiangju Sun, and Xiaowei Cheng

Internet Business Department of State Grid, Gansu Electric Power Company (Information and Communication Company, Data Center), Lanzhou 730050, Gansu, China

Abstract. The rapid economic development makes people's lives inseparable from power communications. In order to ensure people's quality of life, the development and application of smart optical distribution technology, a high-performance power communications technology, has become the current power grid company's intelligent development main direction. As a communication network system that specializes in power grid operation and power supply enterprise management information, the power communication network system is undergoing important technological changes in order to meet the development of power supply at all levels across the country. It is necessary to ensure the smooth transmission of power supply dispatch and enterprise management information. It is also necessary to adapt to the demand for management information transmission of power supply enterprises that has rapidly increased broadband. This article studies the broadband services in a certain province, and analyzes the applicability of smart optical distribution technology in power communication transmission by analyzing the number of broadband users in the province and the usage of broadband traffic in Internet TV.

Keywords: Smart optical distribution technology · Power communication system · Communication technology · Power grid enterprise management

1 Introduction

Traditional power communication mainly relies on SDH technology for service transmission, and the protection of services only relies on its own main-standby switching method. Today, with the vigorous implementation of smart grids, there are major limitations, such as poor network scalability and network transformation. Shortcomings such as insufficient flexibility and insufficient capacity to meet the needs of the surge in business volume are in urgent need of upgrading and transformation. The unique advantages of OTN provide a new direction for the future development of power communication under the smart grid.

Many scholars at home and abroad have conducted research on the application of intelligent optical distribution technology in power communication transmission, and have achieved good research results. For example, the single-wave transmission technology promoted by a power grid company has greatly increased the capacity of the

OTN transmission platform, but for the power communication network, the research on the optical transmission network is more focused on the transmission of long distance between stations, in order to meet the long-distance transmission needs of the UHV grid, research on the transmission performance of SDH systems, including ultra-low-loss optical fiber, super-strong coding error correction, high-power optical amplification technology, etc., but the application research of OTN systems in power communication systems has not been in-depth [1]. A conference put forward the concept of Automatic Switching Optical Network (ASON) technology, conducted in-depth research and discussion on ASON standardization, and put forward some related draft standards or development suggestions accordingly, in order to realize the optical communication equipment of different manufacturers in optical communication Interconnection and intercommunication in the network. The conference standardized ASON's general open structure model, standard interface, and intelligent control protocol for dynamic configuration exchange. It mainly defined the ASON system architecture from the overall network architecture, routing protocol, automatic identification, etc. And protocol specification [2]. Although the research results in the application of smart optical distribution technology in power communication transmission are still good, there are more and more power communication services. To achieve the development goal of smart grid modernization, it is necessary to integrate smart optical distribution technology with power communication services.

This article explains the concept and classification of smart optical distribution technology. The most promising one in the transmission network is OTN technology, and applies OTN technology to the actual power communication transmission system, and analyzes OTN based on the broadband usage of the power grid in a certain province. The applicability of technology in power communication transmission.

2 Intelligent Optical Distribution Technology and Influencing Factors of Power Communication Transmission Performance

2.1 The Concept of Smart Optical Distribution Technology

Intelligent optical distribution technology is based on various transmission technologies, adding an independent control layer to the SDH transmission layer and optical transmission network (OTN), so that signals of different rates can be clearly communicated. Intelligent optical distribution technology can provide stable bandwidth for the optical data transmission channel between the two customer network features at the source point of optical network transmission and the egress point of acceptance. Intelligent optical distribution technology is the development trend of power communication transmission technology in the future. Through the research of intelligent optical distribution networking technology, the flexible and efficient networking problem of the Internet can be solved [3, 4].

2.2 Classification of Smart Optical Distribution Technology

(1) SDH technology

SDH is the multiplexing, demultiplexing, link transmission and exchange of communication services, through the unified network management system for integrated control of synchronous optical network technology, and finally evolved into a synchronous digital system, that is, the SDH communication transmission system. The SDH transmission system sequentially transmits left, right, top, and bottom serial code streams. The basic transmission process is: the sender adds various information code blocks in the information payload in the STM-N frame structure and adds them for channel performance management monitoring, control overhead bytes, and then add network operation management and maintenance (OAM) bytes, and then add a management monocular pointer to ensure effective separation of the information payload and then transmit it to the receiving end [5, 6].

(2) ASON technology

ASON technology supplements automatic switching under routing and signaling control with multi-granularity and multi-level information functions, and provides various personalized services dynamically initiated by users. The new generation of optical networks integrates service applications, automatic routing and signal control to establish or clean up connections, exchange and merge transmission power. The ASON network adopts the MESH networking method to contribute to the improvement of network availability; the automatic detection of network topology, the automatic establishment and deletion of connections, and the end-to-end configuration of services are realized by the introduction of a control plane; through dynamic bandwidth allocation, dynamic configuration and fast generating services effectively enhances the ability to recover from multiple network node failures. The main feature of ASON is that it can provide a fast connection, and the use of distributed control can provide an end-to-end connection within a few seconds. Reducing service delivery time will help operators gain more customers and generate profits [7].

(3) OTN technology

OTN overcomes the shortcomings of the SDH network and is a newly developed transmission technology. It not only combines the technical advantages of SDH, but also achieves high-efficiency and low-cost transmission of large-particle services. With the rapid growth of broadband data services, the use of increasingly mature OTN technology to establish a more efficient and reliable transmission network is an inevitable evolution of OTN technology. The high-speed cross-particle crossover efficiency provided by OTN technology is very high, so that the equipment has a strong cross-connection capability and reduces the cost of purchasing high-performance equipment. Compared with SDH scheduling particles, OTN configuration particles are obviously larger, which greatly improves the adaptation and transmission efficiency of data client services [8].

OTN technology can effectively solve the transmission, planning and protection problems of large-scale data services, provide efficient operation management and maintenance functions for the network, strengthen the functions of the network, and increase

the durability of data services. OTN technology can also be similar to ASON technology. Combine, make the network more intelligent [1].

2.3 Parameters Affecting Transmission Performance

(1) Attenuation

The attenuation of the optical signal in the optical cable represents the energy loss of the optical signal during the transmission process. The energy loss is mainly due to the absorption of the light wave by the material and the reduction of the light wave's own energy during the transmission process. The inherent refinement of the optical fiber spectrum is through the addition of potential absorption loss and scattering loss.

(2) Dispersion

Dispersion is the basic characteristic of optical fiber. The propagation speed of light depends on the refractive index of the medium, and the refractive index of the optical fiber depends on the wavelength of the optical signal. Therefore, the speed of light at different wavelengths in the optical fiber may be different, which is a scattering effect. In optical fiber communication, distortion and expansion conditions will spread when the signal pulse propagates, causing crosstalk between symbols, and ultimately limiting the overall performance of the system. The chromatic dispersion in the optical fiber has two basic effects. One is chromatic dispersion, which is the main problem of optical fiber transmission. The different propagation speeds are caused by the different frequencies of the light waves contained in the signal in the optical cable. The signal passes through the light wave after long-distance transmission. Transmission, causing the front and back phases, the adjacent signal pulses overlap and produce corresponding position errors. The other is polarization mode dispersion, which refers to the irregular circular cross section of the core, and polarization mode dispersion is often the result of local disturbances. The difference in signal propagation speed in the two basic modes of optical fiber leads to different grouping delays [9, 10].

(3) Non-linear effects

There are two types of nonlinear effects in single-mode fibers: stimulated scattering effects and nonlinear refractive index effects. Stimulated scattering is a gain or loss related to light intensity, and nonlinear refractive index causes a phase change related to light intensity [11].

2.4 Calculation of Optical Transmission Section Parameters

The division of the optical multiplexing part must consider the influence of the system signal-to-noise ratio and chromatic dispersion. There are two calculation methods as follows.

(1) Rule design method

In the case that the project length is relatively uniform, a regular design method can be adopted: use the scattering limit formula and the loss limit formula to determine

the signal-to-noise ratio of the system, and then calculate the length of the optical path and take the minimum value.

$$L = \frac{D_{sys}}{|D|} \quad (1)$$

L represents the length of the optical multiplex section with limited dispersion (km), D_{sys} represents the maximum allowable dispersion value (ps/nm) of the optical channel between MPI- S_M and MPI- R_M points, $|D|$ represents the average dispersion coefficient (ps/nm.km).

(2) Simple calculation method of signal-to-noise ratio

When the rule design method does not meet the calculation requirements of the actual application, it can be replaced with a simpler and more convenient SNR calculation method: the length of the optical multiplexing part or the optical transmission part is designed by the system using the scattering limit formula and simple SNR calculation the formula is determined. This method is often used when there is no significant difference in the consumption of the optical transmission section.

$$OSNR_N = 58dBm + P_{out} - 101gM - Nf - Aspan - 101gN \quad (2)$$

$OSNR_N$ represents the optical signal-to-noise ratio (dB) of each channel after N optical transmission sections, M represents the number of channels, P_{out} represents the total incoming fiber power (dBm), Nf represents the noise figure of the optical amplifier, and the maximum optical transmission section loss of $Aspan$ (dB).

3 Application Research of Smart Optical Distribution Technology in Power Communication Transmission

3.1 Research Purpose

OTN technology is an ideal technical system for building large-capacity power communication optical transmission networks. The large-capacity OTN network of the power communication network at all levels is developing rapidly, and the construction projects are relatively concentrated. Therefore, it is necessary to conduct research on the application of OTN technology in power communication systems, and understand the particularity and applicability of the application of OTN key technologies on power optical cable grids. Based on a comprehensive comparison of different transmission technologies, combined with the future technology development direction, the telecommunications network company decided that the new transmission network will adopt OTN technology and build an OTN platform based on large-scale IP broadband services, the business requirements of the metropolitan area network, and optical fiber resources. The contradiction has promoted the further integration and development of the local network transmission network.

3.2 Research Methods

This article collects the number of broadband users and broadband business traffic in 6 cities in a province from 2017 to 2020 in the statistical yearbook, and analyzes the traffic usage of users watching online videos through broadband, proving the application of OTN technology in power communication systems sex.

4 Application Analysis of Smart Optical Distribution Technology in Power Communication Transmission

4.1 Basic Situation of Communication Users

Table 1. Number of users (ten thousand) and metropolitan area network traffic (G)

	2017		2018		2019		2020	
	User number	Broadband	User number	Broadband	User number	Broadband	User number	Broadband
A	3.42	15.23	3.74	18.25	4.18	24.63	4.36	31.79
B	3.56	15.54	3.98	18.67	4.34	26.24	4.72	35.28
C	2.13	11.38	2.37	12.01	2.61	17.51	3.10	23.50
D	1.58	7.34	2.02	8.87	2.57	10.36	2.86	18.67
E	1.26	6.97	1.63	7.85	1.95	9.37	2.44	16.29
F	2.71	12.45	3.15	14.11	3.42	20.48	3.75	28.96

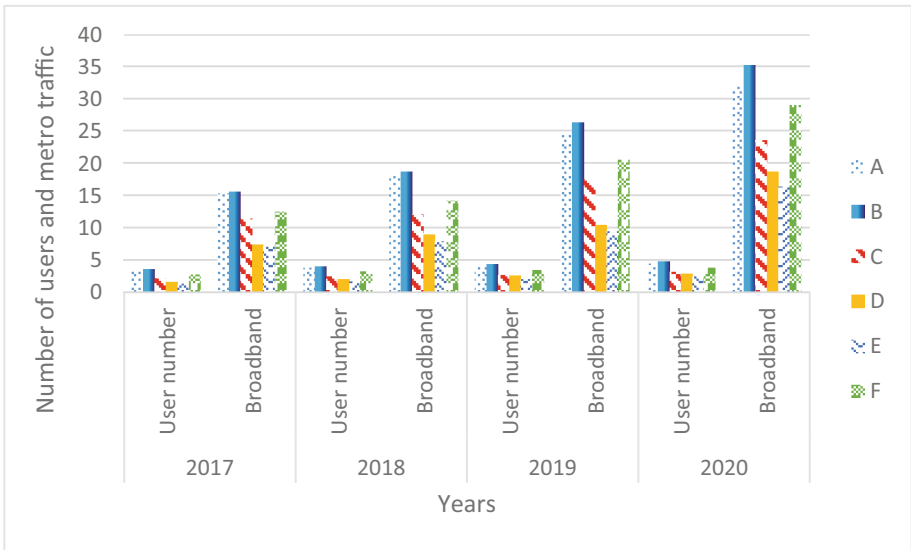


Fig. 1. Number of broadband users and business usage from 2017 to 2020

As shown in Table 1 and Fig. 1, the number of broadband users and broadband usage in 6 cities in a certain province in recent years. From the table, we can see that the number of broadband users and broadband usage in 6 cities has increased year by year, and city A and City B has the largest number of broadband users, with 43,600 and 47,200 by

2020. These two cities are also the cities that use the most broadband traffic, indicating that the province’s communication transmission network coverage is becoming more and more extensive, and the number of people using broadband Will gradually increase.

4.2 Number of Broadband Traffic Used by Internet TV

Figure 2 shows the traffic usage of broadband users in 6 cities using broadband to watch Internet TV. It can be seen directly from the figure that the traffic layer of broadband users in 6 cities using broadband to surf the Internet to watch TV is increasing year by year, and the growth rates of City A, City B, and City F are the fastest, while the growth rates of City D and City E are the fastest. Relatively slow. It shows that the application of OTN technology in the province’s power communication system provides users with communication transmission channels to facilitate users’ access to the Internet.

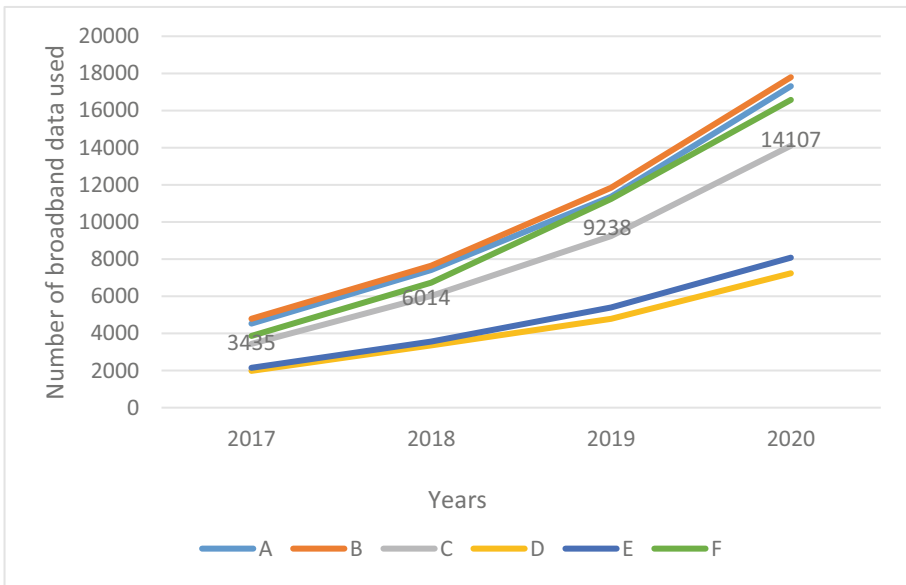


Fig. 2. Internet TV uses broadband traffic (Mbps)

5 Conclusion

The power information communication transmission network is an important infrastructure invested and managed by the state grid to ensure the development of the energy industry. It provides the necessary information transmission and exchange channels for the construction of the power grid, and strongly supports the safe and stable operation of the power grid. This paper studies the application of smart optical distribution technology in power communication transmission, combined with the number of broadband

users and broadband traffic, and finds that OTN technology is more practical in power communication systems. The use of OTN technology can meet user communication requirements. OTN equipment will be the main equipment type of the metropolitan area transmission network in the future networking.

References

1. Farkas, T.D., et al.: Application of power line communication technology in street lighting control. *Int. J. Des. Nature Ecodyn.* **13**(2), 176–186 (2018)
2. Zhang, H., Zheng, B., Zu, H.: Application of SVGA video real-time transmission technology in music education information communication. *J. Intell. Fuzzy Syst.* **39**(10), 1–12 (2020)
3. Sun, Y., et al.: Design of GOOSE communication for IED in intelligent hydropower station. *Dianli Xitong Baohu yu Kongzhi/Power Syst. Protect. Control* **46**(21), 145–151 (2018)
4. Miao, X.J., et al.: Intelligent model and application of power transmission project cost prediction. *Boletin Tecnico/Technical Bull.* **55**(13), 43–47 (2017)
5. Yang, J., et al.: Application of multi-agent technology to fault diagnosis of power distribution systems. *Int. J. Intell. Inf. Technol.* **1**(2), 1–16 (2017)
6. Feng, L., Zhao, J.: The study on the application of the intelligent technology in the sightseeing agricultural parks. *Revista de la Facultad de Ingenieria* **32**(15), 12–17 (2017)
7. Broni-Bediako, E., et al.: Application of intelligent well completion in optimising oil production from oil rim reservoirs. *Adv. Geo-Energy Re.* **3**(4), 343–354 (2019)
8. Kalanidhi, K., Vinodkumar, D., Sridharan, P.: Transmission power line fault detection based on deep learning methods. *Indian J. Power River Valley Dev.* **69**(3–4), 35–38 (2019)
9. Rajendhar, P., Edward, J.B.: Application of demand response and co-simulation approach for renewable integrated home energy management system: a review. *IET Gener. Transm. Distrib.* **13**(16), 3501–3512 (2019)
10. Zhao, Z., et al.: Joint user pairing and power allocation scheme based on transmission mode switching between NOMA-based maximum ratio transmission and MMSE beamforming in downlink MISO systems. *Mobile Inform. Syst.* **2021**, 1–21 (2021). <https://doi.org/10.1155/2021/6671371>
11. Manoj, G., Jacob, E., Kundukulam, S.O.: Relay based coupling scheme of high speed communication data, high voltage DC and high power pulsed AC for coaxial cable. *Def. Sci. J.* **68**(5), 487–493 (2018)



Intelligent Puncture Equipment in Sampling of Gynecological Tumors

Hui Zhu^(✉)

Hebei North University, Zhangjiakou 075000, Hebei, China

Abstract. According to the statistics of the Ministry of Health, the mortality and incidence of breast cancer in Chinese women have been increasing year by year. Research on the key technologies of assisted puncture robots based on medical images can well solve the problems existing in traditional percutaneous puncture operations. The purpose of this article is to study the application of smart puncture equipment in gynecological tumor sampling. Analyze the movement of the puncture needle inside the tissue, and establish a trajectory and planning model based on the robot kinematics. The model uses the puncture coordinates and target coordinates to analyze the needle radius value, which can effectively predict the trajectory of the needle, and at the same time predict the trajectory of a deflection by inputting the coordinates of the inflection point. The experimental results show that the deviation values of the puncture needles are all stable within 8%, which meets the experimental requirements for the stability of the radius value.

Keywords: Smart puncture · Puncture equipment · Gynecological tumor · Tumor sampling

1 Introduction

Puncture cytology can accurately detect breast cancer. It has high sensitivity, simple method, safety, small trauma, low cost, and a wider range of clinical advantages, but it cannot determine the source of cells. Immunohistochemical testing by creating cell wax blocks can further help diagnosis [1, 2]. For patients with high clinical suspicion of malignant tumors, surgical biopsy is recommended [3, 4]. Puncture surgery, also known as invasive medical surgery, refers to a treatment method in which a puncture needle is inserted into human tissue through manual or mechanical surgery for the next step of tissue biopsy or targeted drug therapy. It has been widely used in human biopsy.

There are many scholars who study the whole process of tumor [5]. Some scholars investigate the practice of nurses in solving the sexual problems of patients with gynecological tumors, and find out the contributing factors and obstacles. The results show that the main factors contributing to nurses' sexual health care are good nurse-patient relationship, communication skills and training for nurses. It is necessary to strengthen the training of nurses, increase the number of nurses, and provide patients with all-round care [6]. In addition, scholars have evaluated the routine use of endometrial samples in outpatient clinics through LEVEL trained by gynecological residents. The causes of

endometrial sampling are divided into 6 categories, including screening, PMB, menorrhagia or AUB, huge uterine and cervical lesions. Use Pipelle to obtain an endometrial sample. Endometrial biopsy is accurate, simple, and safe. It can be performed by LEVEL, a gynecological resident, with an accuracy rate of 70%. An additional endometrial examination is required for undiagnosed patients [7]. The application of intelligent puncture equipment in gynecological tumor sampling has practical significance.

This work uses medical images to conduct in-depth research on the basic technology of precision puncture robots, analyze the kinematics and workplace characteristics of the continuous six-degree-of-freedom robot, combine route planning and surgical puncture design, and study the space trajectory of the improved RRT algorithm drilling robot system. The design ensures that the final puncture guide of the precision puncture assist robot can accurately reach the puncture needle point with the best surgical route. Lay the foundation for the clinical application of human breast tumor surgical puncture.

2 Research on Intelligent Puncture Equipment in Gynecological Tumor Sampling

2.1 Accurate Trajectory Planning of the Piercing Robot

After completing the medical image processing, the identification of the marked points of the medical three-dimensional image, and the selection of the puncture path, the medical image coordinate system is registered to the robot surgery space coordinate system through binocular vision, and the auxiliary precision puncture robot system needs to have the best design space. The trajectory planning algorithm enables the end guide of the assisted precision puncture robot to accurately reach the needle point of the puncture operation set by the doctor with the best surgical path [8]. The path planning of the assisted puncture robot must avoid obstacles, select the optimal puncture positioning angle, and be able to move to the best position for the doctor to complete the manual puncture operation.

In medical puncture surgery, with the help of the assisted precision puncture robot system to achieve precise positioning, it is not only necessary to plan the starting point and target point in the surgical space, but also to plan some path points between the two, so that the robot follows a specific Path movement [9]. Through the tool coordinate system to reach the desired position and posture in the working coordinate system, set the path point in the robot motion, and use the researched robot inverse kinematics principle to convert the position and posture corresponding to the path point into the corresponding joint angle [10, 11].

In order to realize the trajectory planning of the joint space of a multi-degree-of-freedom series robot, a high-order polynomial is usually used to solve the position, the speed and acceleration of the joint motion, etc. The net is sent to the robot controller, and controls the robot to do the corresponding movement [12].

2.2 Dynamic Tracking Measurement Principle Based on Position Vision

The principle of dynamic tracking based on binocular position vision to measure the micro-movement of the patient's body mainly includes camera calibration, image acquisition and calibration. Recognition and extraction of marked points, stereo matching,

calculation of three-dimensional coordinates, extraction of marked points after body movement, matching, and acquisition of new three-dimensional information. The design scheme of dynamic tracking and measuring marker points based on position vision mainly completes the following process: the binocular camera quickly collects patient images and analyzes and processes them, recognizes and stereo matches the marker points, and derives the position of the marker points in the camera coordinate system. Spatial three-dimensional coordinate information; when the patient's body is slightly moved, its circular marking point deviates from the original position, the binocular camera collects image information again and analyzes and processes it, and after identifying and stereo matching the marking point, the circular marking point is deduced again. The new spatial three-dimensional information in the camera coordinate system; the spatial three-dimensional coordinate difference of the circular mark point is used to solve the change of the corresponding spatial position and posture of the mark point after the micro-movement occurs, and it is done in the process of assisting the precise positioning of the puncture robot. Error compensation enables the puncture guide tube to be accurately positioned on the puncture needle point after the patient's body is slightly moved.

3 Research and Investigation of Intelligent Puncture Equipment in Sampling of Gynecological Tumors

3.1 Construction of a Robotic Arm Assisted Puncture Experiment Platform

Fix the UR3 six-degree-of-freedom mechanical arm to the OTP damping isolation platform, and install a specially designed clamp to hold the puncture needle at the end of the tool. The puncture needle used in the three-dimensional path planning verification experiment based on the improved RRT algorithm is the No. 2 puncture needle with a diameter of 0.8 mm and a length of 80 mm. Subsequent PVA prosthesis and biopsy puncture comparison experiments were performed with 20G, 21G, 22G, 23G puncture needles, the puncture needle was installed on the fixture, and the PVA prosthesis was placed on the plate. Fix the camera at a distance of 20 cm from the PVA prosthesis.

3.2 General Information

Collected patients who were planned to undergo needle biopsy of breast lesions in the University Hospital from January 2021 to October 2021. Inclusion criteria: the maximum diameter of breast lesions is greater than 5mm; imaging assessment BI-RADS 4a or higher. Exclusion criteria: Pathological diagnosis has been made before needle biopsy; there is no obvious solid part in the ultrasound appearance of the lesion. Finally, this study selected 100 patients who underwent ultrasound-guided needle biopsy, including 5 males and 95 females. The age of the patients for the first biopsy was 30–50 years old, with an average age of 40 years.

3.3 Breast Tumor Puncture Biopsy Method

Preoperative Preparation: The patient undergoes routine electrocardiogram examination before the operation to ensure that the patient has no serious heart disease, and blood is drawn for routine blood tests and coagulation to ensure that the total number of platelets and PT time are within the normal range. If the patient takes aspirin and other anticoagulant drugs before surgery, they should stop taking it, and perform a puncture operation after checking that the blood coagulation is normal one week later. Sign the informed consent form for the operation and the consent form for the use of high-value medical consumables.

Intraoperative Operation: Before the operation, the doctor performed ultrasound again to determine the location of the nodule and the course of the adjacent blood vessels, and determine the route of needle insertion to avoid complications such as pneumothorax. The operating physician wears sterile gloves, the needle tip reaches the breast mass, and pulls the trigger to release the needle groove. Generally take 3–4 tissue strips about 2 cm in length and fix them with formalin solution for inspection. After removing the needle, cover the puncture point with a small applicator.

Postoperative Precautions: pressure the needle track at the gauze layout for 30 min to prevent bleeding.

3.4 Data Preprocessing

The force sensor can be used to measure the magnitude of the resultant lateral force. Therefore, the following needle tip bending deformation estimation equation and bending radius calculation formula are established:

$$bend_B = -\frac{F_V l^4}{8Elh} \quad (1)$$

Among them, l represents the length of the flexible needle (m), which can be obtained by measurement. E represents the elastic modulus of the needle material, and I represents the moment of inertia of the cross-sectional area.

In addition, whether the puncture probe can accurately reach the center of the lesion according to the planned path is also a very important evaluation indicator. Therefore, the three indicators of the number of rotations n , the total length of the path L and the distance d from the end of the path to the center of the lesion are taken into consideration together. Different weight ratios w_1 , w_2 , and w_3 are assigned to these three, and the calculation formula of the new inspection index is shown in 2.

$$W = w_1 n + w_2(L - D) + w_3 d \quad (2)$$

4 Research and Analysis of Intelligent Puncture Equipment in Sampling of Gynecological Tumors

4.1 Results and Analysis of Three-Dimensional Path Planning Based on the Improved RRT Algorithm

For a flexible puncture needle with a diameter of 0.8 mm and a needle length of 80 mm, the bending radius of the flexible puncture needle can be calculated to be about 300 mm.

According to the three-dimensional path planning process of the puncture probe, the process of puncturing such a flexible puncture needle into the breast tumor tissue is planned, and we can get 10 puncture paths of the flexible needle. The arc length L and the number of rotation segments n of these 10 puncture paths are shown in Table 1 (Fig. 1).

Table 1. Puncture path data of 10 flexible needles

Serial number	Number of rotations	Path length (mm)	W
1	4	228.12	2.42
2	4	228.63	2.31
3	5	228.24	2.5
4	5	228.68	2.27
5	5	228.45	2.15
6	4	228.44	2.1
7	5	228.31	2.34
8	4	228.29	2.33
9	6	228.76	2.51
10	5	228.68	2.22

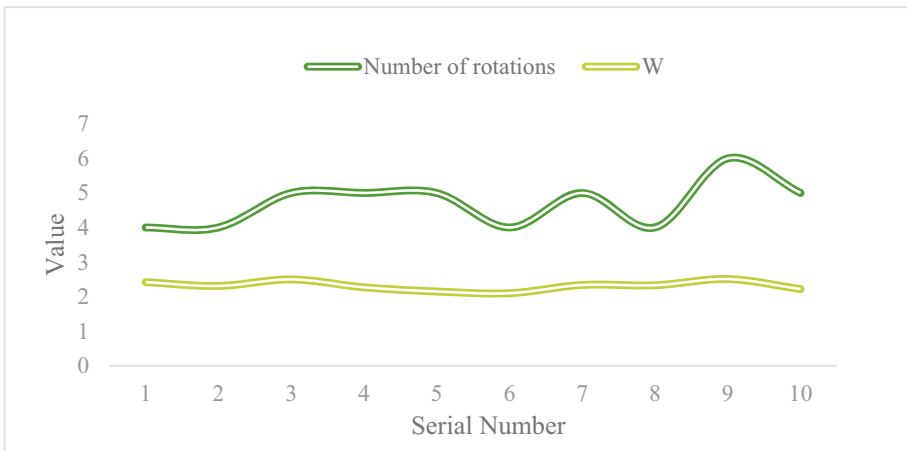


Fig. 1. Puncture path data of 10 flexible needles

The linear distance D from the needle insertion point to the center of the lesion is 226.6 mm. Using the method of optimal path solving, the three evaluation indicators are respectively 0.3, 0.4 and 0.5. We can calculate the optimal puncture path among these 10

puncture paths as the sixth puncture path obtained by simulation. The trajectory length of this puncture path is 228.44 mm, and it has been rotated 4 times in total.

According to the planned puncture path, the puncture needle can avoid obstacles well and finally reach the target lesion. At the same time, the planned puncture path does not have a large reentry. It is an arc path, which fully conforms to the physical characteristics of the puncture needle.

4.2 PVA Prosthesis and Biopsy Comparison

During the needle puncture process, due to external interference and the imaging principle of the instrument itself, the noise in the collected images will also be slightly different, resulting in different image quality exported by the instrument. To explore whether these noises will affect the results of the three-dimensional path planning of the improved RRT algorithm. Using 20G, 21G, 22G, 23G puncture needles for puncture, due to the unevenness of the PVA prosthesis, the bending radius of the needle body of the puncture needle fluctuates within a certain range, but the upper and lower deviations of the 4 types of puncture needles fluctuate. It is less than 150 mm, showing good stability.

When performing biopsy puncture, the dynamic tracking measurement of binocular position vision is used for image acquisition. During puncture, the puncture needle bends downward, and the collection surface always overlaps with the curved surface of the puncture needle. The puncture experiment is divided into two types: core-pulling and non-core-pulling, both of which are non-deflection puncture experiments. In the experiment, four types of puncture needles are used for puncture. The deflection radius of the puncture needle in the biopsy is shown in Fig. 2.

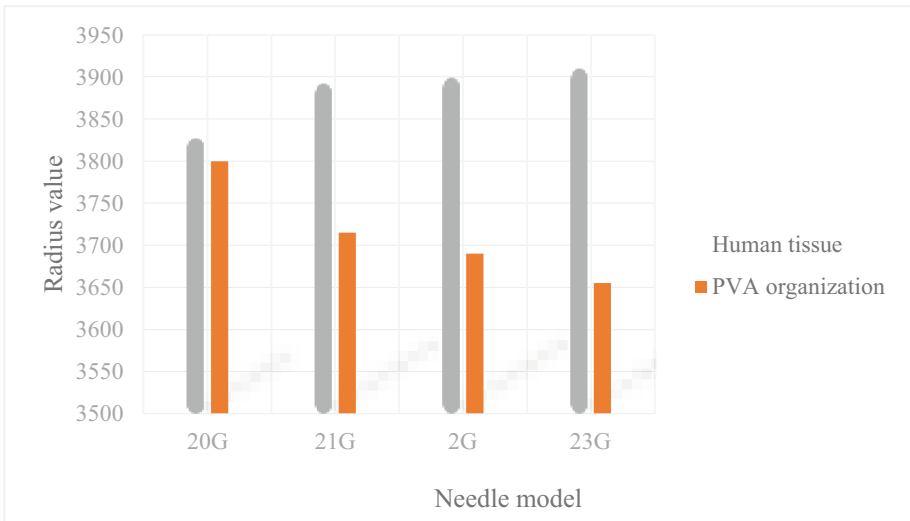


Fig. 2. Comparison of PVA prosthesis and biopsy

The total radius value difference is stable within 8%, indicating good stability. The radius of the needle biopsy is larger than that of the PVA prosthesis. This is because the tissue structure of the human body is looser than the PVA prosthesis, which makes the cutting action on the needle injection smaller than the force in the PVA prosthesis. In addition, the internal structure of muscle tissue is more complex than that of PVA tissue prosthesis, and has more uncertainty, making the trajectory of the puncture needle less stable in some cases. The release radius of the puncture needle with and without core pulling differs by 7%, indicating that the puncture needle has the best flexibility when removing the core. With the increase of the puncture needle model, the radius of the puncture needle deflection in the tissue is also decreasing, and the curvature of the puncture needle shows a certain linear relationship. The deviation values are all stable within 8%, which meets the test requirements for the stability of the radius value.

5 Conclusions

In order to improve the cure rate of patients, the treatment of breast cancer has changed from anatomical mode. However, before clinical diagnosis and treatment of breast cancer, a biopsy must be performed to obtain a baseline medical history before further specific treatment can be performed. This paper discusses the position of the robot puncture system based on binocular vision, and proposes a mapping algorithm from the medical imaging system to the robot projection coordinate system to solve the problem of system field registration in puncture. Robotic puncture assists imaging from site to surgical site. It can provide doctors and patients with more thinking time and choose personalized treatment plans according to the situation, so that the treatment can be more effective in the treatment of breast cancer.

References

1. Kalshetty, A., Basu, S.: Non-18F-2-fluoro-2-deoxy-d-glucose PET/computed tomography in gynecologic oncology: an overview of current status and future potential. *Pet Clin.* **13**(2), 239–248 (2018)
2. Aivazova-Fuchs, V., Mierina, I., Friese, K.: *Gynakologe* **51**(12), 1060–1063 (2018). <https://doi.org/10.1007/s00129-018-4302-9>
3. Annunziata, C.M., Kohn, E.C.: Clinical trials in gynecologic oncology: past, present, and future – ScienceDirect. *Gynecol. Oncol.* **148**(2), 393–402 (2018)
4. Yang, W., et al.: Establishment of the PDTX model of gynecological tumors. *Am. J. Transl. Res.* **11**(6), 3779–3789 (2019)
5. Liu, K., et al.: Potential value of circulating tumor DNA in gynecological tumors. *Am. J. Transl. Res.* **12**(7), 3225–3233 (2020)
6. Kameyama, H., et al.: Digestive surgery intervention for gynecological malignant tumor. *Gan to kagaku ryoho. Cancer & chemotherapy*, **46**(13), 2176–2178 (2019)
7. Over, D.R., Rust, S., Fendley, H.F.: Venous thromboembolism prophylaxis in postoperative gynecologic oncology patients. *J. Ark. Med. Soc.* **113**(11), 266–268 (2017)
8. Lefkowitz, C., et al.: Opioid use in gynecologic oncology in the age of the opioid epidemic: Part I – Effective opioid use across clinical settings, a society of gynecologic oncology evidence-based review. *Gynecol. Oncol.*, **149**(2), 394–400 (2018)

9. Jeske, Y.W., et al.: FGFR2 mutations are associated with poor outcomes in endometrioid endometrial cancer: an NRG Oncology/Gynecologic Oncology Group study. *Gynecol. Oncol.*, **145**(2), 366–373 (2017)
10. Reiser, E., et al.: Impact of frailty on the management of patients with gynecological cancer aged 80 years and older. *Arch. Gynecol. Obstet.* **303**(1), 1–7 (2021)
11. Liu, S.N., et al.: Risk factors analysis of venous thromboembolism in post-operative patients with gynecological malignant tumor and application of related risk assessment table. *Acad. J. Second Mil. Univ.* **38**(10), 1244–1249 (2017)
12. Obermair, A., et al.: Nutrition interventions in patients with gynecological cancers requiring surgery. *Gynecol. Oncol.*, **145**(1), 192–199 (2017)



Design of Environmental Art Optimization System Based on Improved Particle Swarm Optimization Algorithm

Yi Fu¹ (✉), Chunjiang Yu¹, and Saad Metawa²

¹ School of Design and Art, Shenyang Jianzhu University, Shenyang 110000, Liaoning, China

² Mansoura University, Mansoura, Egypt

Abstract. With the enhancement of environmental awareness, people pay more and more attention to environmental art design. Starting with the improved particle swarm optimization (PSO) algorithm, this paper proposes an environmental art system based on improved PSO. The improved PSO algorithm is applied to the environmental art system to solve the problems faced by the current environmental art. Firstly, this paper introduces the connotation and characteristics of environmental art system, and then studies the application of improved PSO algorithm. According to the system design principles, the environmental art system is designed and developed, and its performance is tested. Finally, the test results show that the four modules of art painting, architectural appearance, interior design and landscape design of the system have good performance in the test, and the scores are more than 3, which shows that the system can meet the operation requirements.

Keywords: Improved particle swarm optimization · Environmental art · System design · System optimization

1 Introduction

In recent years, with the continuous development of Internet technology, environmental art design is gradually developing towards a more diversified and multi-functional direction. Nowadays, human beings pay more and more attention to the ecological environment. In today's society, people are not only satisfied with the visual aesthetic function of the environment in the traditional sense [1, 2].

Many scholars have studied the improved PSO algorithm. Some scholars have improved PSO algorithm by introducing global optimal selection pool and probabilistic global optimal selection pool (p-gbest). The algorithm obtains a set of Pareto optimal (PO) solutions. By calculating the fuzzy membership (FM) degree of Pareto optimal solution set, the individual with the smallest FM degree is selected as the final optimization result. Through the comparative analysis of the results, it is proved that the optimization effect of the algorithm is excellent [3, 4]. Some scholars have proposed a solution based on the improved multi-objective PSO algorithm. Firstly, a multi-objective decision-making model is established according to the delay, area and power of large-scale mprmm circuit.

Then, on the MCNC benchmark of PLA format, the proposed solution is compared with the current three preferred algorithms, and the results verify the effectiveness of the solution [5, 6]. The above research lays a foundation for the design and research of environmental art optimization system based on improved PSO algorithm.

In this paper, a method based on Improved PSO and fuzzy clustering evaluation system is proposed to realize the real-time monitoring of pollutant emissions in urban ecological environment, social economy and other fields. By designing a comprehensive system and integrating it with traditional artificial neural network technology, the purpose of improving global performance and reducing computing cost is achieved. This method can make environmental art have good adaptability and reduce the problem of resource waste or pollution caused by human factors to a certain extent.

2 Discussion on Environmental Art Optimization System Based on Improved PSO Algorithm

2.1 Environmental Art System

2.1.1 Connotation

In the early 1990s, Chinese local designers had a superficial understanding of the concept of environmental art system design. They consciously and subconsciously engaged in design and made profits, which made up for the lack of the development of Chinese architecture. In the 21st century, our current design theme is system design [7, 8]. Environmental art design includes not only interior design and landscaping, but also system design, just as foreign villa design includes the appearance, landscape and interior of the villa. In the United States, they call it the trinity of planning, architecture and the overall landscape. In the systematic design of environmental art, painting, architectural appearance, interior, landscape and orientation, all design elements are combined with ideas to form an overall design behavior system. For example, the overall architectural design is a new architectural design concept, which is completely synchronized with various components, external environment, building structure, technical equipment and so on. Integrating new technologies means using the technical knowledge of various disciplines to develop and apply new technologies to adapt to external space, building volume and building components, so as to reduce investment and cost [9, 10].

2.1.2 Features

China's environmental art design must strive for simplicity, freedom and full control in order to create high-quality environmental art [11, 12]. Because it promotes human beings in the process of feeling art, not limited to the intuition of the ear, but to mobilize the imagination of the soul, so as to capture the image of environmental art to a higher level. In the historical period when the development of human civilization is still at a low level, we must further put forward requirements for natural environment art, so as to put more emphasis on transformation and conquest. When the development of architectural civilization reaches a higher level, the harmonious relationship between architecture and nature will come first. Environmental aesthetics refers to the trinity of thinking, emotion

and science and technology. When paying attention to the overall architectural design, architects also need to make a breakthrough in creative ideas and design methods. The outbreak of creative inspiration will rely more on the analysis, synthesis and evaluation of the overall design of buildings. We need to learn and improve the overall ability of architecture and overall design, new knowledge system and new way of thinking, as well as the integration of multi-disciplinary technology. We need to use new design concepts to highlight the integrity of architectural creation. The concept of comprehensive building planning can not only improve the adaptability of buildings, but also be the inevitable trend of environmental protection and energy conservation consciousness. With the continuous development of China's economy, science and technology, more and more advanced technologies and scientific and technological achievements are applied to architecture. It is firmly believed that in the near future, there will be more excellent overall architectural design in China's architectural industry. We should constantly sum up and move forward, which is a major responsibility entrusted to architects and designers by history.

2.2 Improved PSO

2.2.1 Algorithm Development

Social creatures living in the wild can reproduce and survive in the wild because of their special abilities. After studying the relevant laws, people developed an evolutionary algorithm corresponding to the group lifestyle and applied it to the related optimization problems. In the bird pile model, anyone can intuitively simulate the bird pile phenomenon by following the following three rules:

- (1) Avoid collision with nearby people;
- (2) Airspeed corresponds to adjacent individuals;
- (3) Aim at the person in the best position;

By studying the behavior of birds, the PSO model is inspired by the foraging law of birds. The space is chaotic and orderly to obtain the optimal solution.

Inspired by bird predation, PSO algorithm has evolved into an artificial intelligence algorithm to solve optimization problems. For example, suppose a group of birds are randomly looking for the only piece of food in the spatial area. Although birds may not know where the food is at first, they can sense the distance between the food and themselves. Therefore, the best strategy for birds is to look for food near the birds closest to the food. In the case of optimization problem, food is the optimal solution of the problem, and the distance between bird and food is the sufficiency of the function. The process of bird looking for food is the best point of the process of function looking for food.

2.2.2 Calculation Process

Based on the illumination in the bird swarm phenomenon, the solution to all optimization problems is that each bird finds food in the d -dimensional search space, which is called "particle". The speed can be adjusted according to the experience of individuals and

populations. When the algorithm initializes the population in the search space, it randomly assigns the position and velocity to a random particle and generates the original population. Then, each particle looks for the optimal solution to the individual or group problem in the complex airspace according to the flight law. In the optimization process, the particle follows two extreme values and updates its state. Extremum is the optimal solution (gbest) found by the whole population, which is called global extremum. The other extreme value is the optimal solution (particle optimization, abbreviated as pbest), which is found by the particle itself, which is called individual extreme value. Particles follow individual extremum and total extremum, which can more completely record the position of the optimal solution.

$$v_{id}(k+1) = v_{id}(k) + c_1 \text{rand}(p_{id}(k) - x_{id}(k)) + c_2 \text{rand}(p_{gd}(k) - x_{id}(k)) \quad (1)$$

$$x_{id}(k+1) = x_{id}(k) + v_{id}(k+1) \quad (2)$$

Where $I = 1, 2, \dots, m$, M is the total number of particles representing m solutions in the group; $D = 1, 2, \dots, n$, n is the dimension of particle search space; $X_{id}(k)$ is the d -dimensional component of the k -th generation particle I ; $V_{id}(k)$ is the d -dimensional component of the airspeed vector of the iterative particle k -th I ; $P_{id}(k)$ is the d -dimensional component of particle I in the k -th iteration. $P_{gd}(k)$ is the best point for the k -generation population to find the component of the d -Dimension itself; c_1 and c_2 are generally a constant set 2. Appropriate c_1 and c_2 can accelerate the convergence of the algorithm and make it more difficult for particles to fall into local optimization. $\text{Rand}()$ is a random number between 0 and 1.

3 Experiment

3.1 System Design Principles

3.1.1 Principle of Continuity and Foresight

The system has the overall concept of overall system architecture, data architecture, physical architecture, security architecture, application architecture and specification. Based on the expansion and increase of functions, corresponding further development and improvement shall be carried out to ensure the sustainable construction of the project. The overall design of the platform should be advanced in technology and involve the design of the system framework and software framework.

3.1.2 Advanced and Economical Principles

The platform adopts advanced technology in line with the development trend of information technology, and adopts advanced technology and architecture model related to this field as far as possible. The system life cycle is long. Try to choose the latest technology. The environmental art data analysis platform adopts Hadoop data processing technology, big data spark technology, spring MVC framework technology, redis technology and other related advanced technologies. The structure of the platform must take into account the principle of economy and make rational use of existing hardware, software, security, network and other resources.

3.1.3 Stability and Safety Principles

Stability is an important principle when designing a platform. In order to ensure the normal operation of the platform, the environmental art data analysis platform faces many users and needs to meet the daily peak needs of the system. At the same time, it must also adapt to special situations such as one-time batch processing, processing a large amount of data and sudden increase of business. The platform shall be established on the basis of mature and stable hardware and application software environment, prevent damage, modification and leakage by improving various security measures, and protect platform hardware, software and data from accidental and malicious causes.

3.2 Environmental Art System Design

Figure 1 is the design structure of environmental art system.

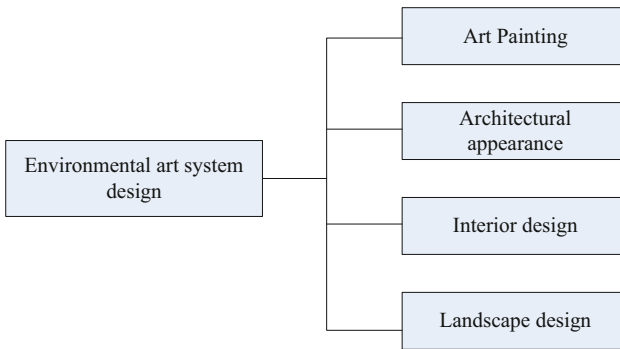


Fig. 1. Structural diagram of the environmental art system

As can be seen from Fig. 1, the environmental art system has the following four parts: art painting, architectural appearance, interior design and landscape design. From painting, architectural appearance, interior, landscape to orientation, all elements are integrated with each other. The system can enable designers to build the source of creative inspiration on the comprehensive ability of system design, and greatly improve the efficiency of environmental art design.

4 Discussion

4.1 Performance Test of Environmental Art System

This paper makes a series of tests on the performance of environmental art system, including load test, capacity test, strength test and pressure test. This experiment adopts the score system. 5 is the full score. The closer the score is to 5, the better the performance of the module. It can be seen from Table 1 and Fig. 2 that the four modules of art painting, architectural appearance, interior design and landscape design of the system have good performance in the test, with scores above 3, which shows that the system can meet the operation requirements.

Table 1. Performance testing

	Load test	Volume test	Strength test	Pressure test
Art painting module	4	5	4	4
Building appearance module	3	5	3	3
Interior Design Module	3	5	4	2
Landscape Design Module	4	5	4	3

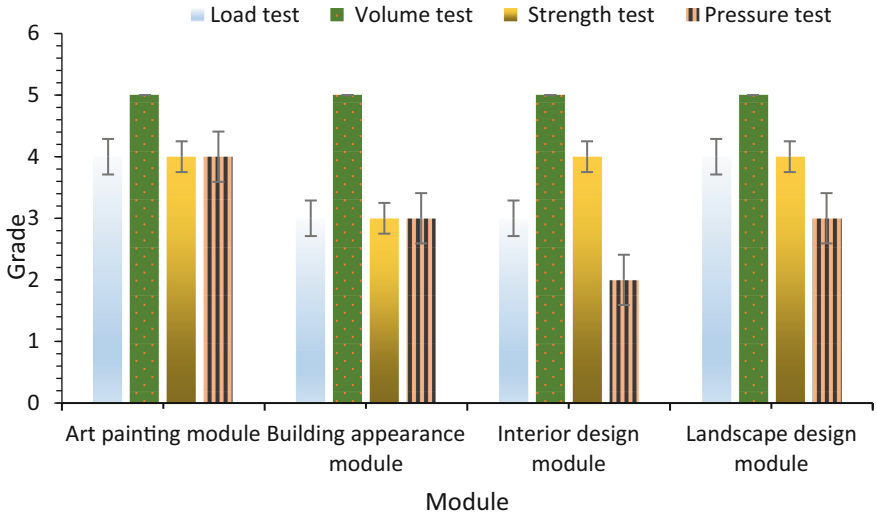


Fig. 2. Performance comparison

5 Conclusion

Environmental art is a science based on ecological environment, which is closely related to environmental protection. With the increasing awareness of environmental protection and ecological protection, people begin to know more about green and clean energy. In this paper, a performance optimization design method of environmental protection system based on Improved PSO (PSO) is proposed, which is applied to the ant colony behavior theory to establish a new model and calculate and solve it. Simulation experiments

verify the feasibility, practicability and effectiveness of the scheme in environmental art, so as to provide users with a better reference and decision support.

References

1. Jalal, M., Mukhopadhyay, A.K., Grasley, Z.: Design, manufacturing, and structural optimization of a composite float using particle swarm optimization and genetic algorithm. *Proc. Inst. Mech. Eng. Part L. J. Mater. Design Appl.* **233**(7), 1404–1418 (2019)
2. Shanmugasundaram, K., Mohmed, A., Ruhaiyem, N.: Hybrid improved bacterial swarm optimization algorithm in hand-based multimodal biometric authentication system. *J. Inform. Commun. Technol.* **18**(2), 123–141 (2019)
3. Al-Qazzaz, N.K., Sabir, M.K., Ali, S.H.B.M., Ahmad, S.A., Grammer, K.: Multichannel optimization with hybrid spectral-entropy markers for gender identification enhancement of emotional-based EEGs. *IEEE Access* **9**, 107059–107078 (2021)
4. Sharifipour, M., Nakhaee, A., Yousefzadeh, R., Gohari, M.: Well placement optimization using shuffled frog leaping algorithm. *Comput. Geosci.* **25**(6), 1939–1956 (2021)
5. Orlando, C., Ricciardello, A.: Analytic solution of the continuous particle swarm optimization problem. *Optim. Lett.* **15**(6), 2005–2015 (2020)
6. Kim, S.-H., Hong, Y.-D.: Dynamic bipedal walking using real-time optimization of center of mass motion and capture point-based stability controller. *J. Intell. Rob. Syst.* **103**(4), 1–17 (2021)
7. Zaharis, Z.D., Gravas, I.P., Yioultis, T.V., et al.: Exponential log-periodic antenna design using improved particle swarm optimization with velocity mutation. *IEEE Trans. Magn.* **53**(6), 1–4 (2017)
8. Dileep, G., Singh, S.N.: An improved particle swarm optimization based maximum power point tracking algorithm for PV system operating under partial shading conditions. *Solar Energy* **158**, 1006–1015 (2017)
9. Xin, Z., Man, Z., Weizhou, W., Yang, J., Jing, T., et al.: Scheduling optimization for rural micro energy grid multi-energy flow based on improved crossbreeding particle swarm algorithm. *Trans. Chinese Soc. Agric. Eng.* **33**(11), 157–164 (2017)
10. Wang, C.F., Liu, K.: An improved PSO algorithm based on comparative judgment. *Nat. Comput.* **17**(3), 641–661 (2018)
11. Chen, C., Li, C.: Process synthesis and design problems based on a global particle swarm optimization algorithm. *IEEE Access* **9**, 7723–7731 (2021)
12. Charfi, S., Atieh, A., Chaabene, M.: Optimal sizing of a hybrid solar energy system using particle swarm optimization algorithm based on cost and pollution criteria. *Environ. Progress* **38**(3), 13055.1-13055.10 (2019)



Software-Assisted Japanese Phonetic Teaching Based on Phonetic Visualization

Xinyi Zhang¹(✉) and Lin Liu²

¹ South China Normal University, Guangzhou City 510631, Guangdong, China

² Shanghai Provident Fund Management Center, Shanghai 201600, China

Abstract. Phonetics teaching plays an important role in Japanese professional courses. Nowadays, various voice-assisted teaching application software are available, but how to make full use of such resources and skills to closely integrate with actual teaching activities and improve students' current voice teaching methods is still a difficult point. This article aims to study the software-assisted Japanese speech teaching based on speech visualization. Based on the analysis of the advantages of speech visualization in speech teaching, the characteristics of Japanese, the reasons for the difficulty of learning Japanese, and the strategies of Japanese speech teaching, it is necessary to understand the advantages of speech visualization software. The application effect of Japanese phonetics teaching, this article conducted a three-month experimental teaching for students in a class of Japanese major in a university, and selected another class with equivalent Japanese proficiency as the control class for comparison. The experimental results show that the software-assisted Japanese speech teaching based on speech visualization technology is effective and feasible.

Keywords: Voice visualization · Software-assisted teaching · Voice teaching · Japanese teaching

1 Introduction

Phonetic practice is the most basic and important part of Japanese teaching. Mastering accurate Japanese pronunciation and tones is a key factor in successfully completing the oral communication tasks in Japanese [1, 2]. However, many Japanese learners in our country are deeply influenced by the local dialect, so sometimes their pronunciation is not satisfactory. Once the speaker's pronunciation is wrong or cannot be mastered by the reader, no matter how good their language level and vocabulary comprehension ability is, normal and efficient oral communication cannot be realized at all [3, 4].

Although the auxiliary teaching activities of multimedia teaching equipment have been widely used in the classroom, the focus is still on listening, reading, and writing. Some teachers still use traditional listening and speaking methods in traditional phonetic teaching, so they cannot effectively combine modern educational information technology with traditional phonetic teaching [5, 6]. In the 1990s, the theoretical support of modern educational information technology also made new progress, and a multi-modal speech

theory based on the operating mode of the operating system emerged. Multimodal speech theory also supports that learners can mix multiple sensory experiences in the process of speech teaching, such as using images to evoke visual experience, or at the same time provide sound effects that evoke the learner's audio-visual experience [7, 8]. The various sensory experiences mentioned above can greatly mobilize students' voice learning motivation. Therefore, speech visualization is also another main technical realization form supported by multimodal speech theory. In recent years, due to the continuous advancement of computer technology, voice visualization tools and software systems have been produced. Such software has the functions of recording, storage, playback, voice fusion, big data analysis, etc., providing powerful capabilities for the management of various voice materials [9, 10]. Through these software systems, abstract audio files can be converted into sound maps with certain voice features, and voice classification can be achieved by extracting the features of related voice data. Among them, Praat speech visualization software is widely used by speech researchers to analyze, mark, edit and synthesize digital speech signals to create various speech maps and text reports, but it has not been widely used in speech education [11, 12]. However, many scholars have conducted research on the education of speech fragmentation characteristics with the help of Praat software, and they have shown that it can optimize the effect of speech education.

On the basis of consulting a large number of references related to speech visualization and speech teaching, this article combines the advantages of speech visualization in speech teaching, the characteristics of Japanese, the reasons for the difficulty of learning Japanese and the strategies of Japanese speech teaching, and selects a Japanese major in a university of the two classes, one class is used as an experimental class that uses voice visualization software to assist Japanese voice teaching, and the other class is used as a control class for teaching in a traditional mode. Compare the differences in attitudes and performance of the two classes on Japanese voice. It is proved that the voice visualization software is effective in assisting Japanese voice teaching.

2 Research on Software-Assisted Japanese Phonetic Teaching Based on Phonetic Visualization

2.1 Advantages of Speech Visualization in Speech Teaching

- (1) It is difficult to reveal the hidden features of speech only by hearing. Voice visualization technology can represent tacit knowledge. The education planning link also directly provides a visual interface, such as the related classification and dynamic display of sound waveform, fundamental frequency, energy, and audio data. The sound map, also called the "holographic photo" of sound, can freeze the sound and re-expand it in a certain time sequence, and help us see the dynamic changes of the sound. And these intuitive sound analysis will be closely linked with the basic plan and final realization of teaching to enhance the effectiveness of education and learning.
- (2) In the process of speech education, different students' initial knowledge level, cognitive strategies, and ability to construct knowledge are unequal, and the speed of

knowledge acquisition also varies. The visual voice learning form allows students to repeatedly see difficult points and enhancement areas regardless of time or field, choose learning methods, and achieve self-learning knowledge and information goals.

2.2 Features of Japanese

(1) Vocabulary

Due to historical factors, the origins of Japanese words are also complicated. First of all, there are several Japanese vocabulary that are specific to Japanese. And they are generally called “Japanese” words in phonetics. This kind of word is now generally spelled in hiragana or Chinese characters, and the pronunciation characteristics of Chinese characters often appear in the form of education. Secondly, due to the long-term influence of ancient Chinese literature, there are Japanese Chinese researchers who directly imported from China and modern Japanese phonetics researchers based on the characteristics of Chinese morphology. These words are generally written directly in Chinese, and the pronunciation of Chinese is also the phonetic pronunciation. The pronunciation characteristics of these words are imported from Chinese other than ancient Chinese, especially “Japanese”, which has been absorbed by Japanese. Today, they are usually written in katakana and are called “foreign words”. Among them, foreign words are the main component of modern Japanese words. But it is worth mentioning that foreign words are not limited to vocabulary introduced in Chinese other than Chinese, but also Chinese vocabulary introduced from postmodern Japanese. With the large-scale adoption of the above three words by the Japanese, some new Japanese vocabulary has gradually emerged, the composition of which is a mixture of two or more of the above three words. This kind of vocabulary is generally called “hybrid”. One of the most common forms of expression is: Chinese character vocabulary with more than two words, the pronunciation method of Chinese characters is often more than one, or the pronunciation and pronunciation of the characters appear together.

(2) Grammar

As a cohesive speech, the characteristic of Japanese grammatical form is between isolated speech and gradient speech. Part-of-speech words will not be distorted due to changes in different syntactic elements. On the contrary, accusation words generally change according to their meaning in the sentence pattern. After adding virtual characters that can represent syntactic elements, they are divided into stems and curvilinear words. Except for some irregular changes, modern Japanese distorted predicate words can only express various syntactic meanings simply by modifying the endings. Unlike most other languages in the world, adjectives contain both verbs and adjectives. This is equivalent to the adjective types in other Chinese. Japanese is divided into “adjectives” and “adjective verbs”. In addition, because Japanese nouns do not change with different names, there are no comparative or superlative adjectives. From the perspective of tense, Japanese predicates also have the distinction between past tense and non-past tense.

2.3 Reasons for Difficulty in Learning Japanese

(1) Negative migration

As a category of cognitive psychology, language communication refers to the process in which language learners transfer existing cognitive experience and experience to language learning and gradually master new things in the process of gradually mastering new things. It can be divided into positive migration and negative migration. If the old knowledge can promote the mastery of the new knowledge, it is called positive transfer, and if the existing experience can hinder the mastery of the new knowledge, it is called negative transfer. In the process of learning Japanese, it is easy to think that the difference between the native language and the target pronunciation is directly proportional to the difficulty of the native language. In fact, the difficulty of the native language is the same, so there is no most difficult. Although there are not many phonemes in Japanese, it is very difficult to master because of the negative transfer of Chinese. Judging from the many difficulties in Japanese pronunciation introduced in the previous article, part of the reason is that due to the lack of similar pronunciation in Chinese, students deliberately emphasize prominent positions during practice. Therefore, learners are more likely to adopt Chinese pronunciation or English pronunciation acquired earlier than Japanese, rather than the most accurate Japanese pronunciation.

(2) Psychological factors

In the process of learning Japanese, in addition to the negative transfer of the mother tongue, there are also psychological factors. Some learners have poor Japanese proficiency, so they also have the intention to learn before learning, but due to the increase in content and learning difficulties, they often give up halfway. However, these learners have serious psychological barriers and are afraid of making mistakes in pronunciation. Therefore, they dare not communicate directly with Japanese people and do not have the necessary daily training. As a result, there is no effective correction of incorrect pronunciation, so that they develop long-term Mispronunciation habits.

2.4 Strategies of Japanese Phonetics Teaching

(1) Grasp the difficulties and break through each

In the voice education stage, the teacher pays attention to observing and analyzing the learners' problems in voice acquisition, so that the learners can circumvent them reasonably, so as to provide appropriate guidance and important training in a timely manner. Practice can correct pronunciation errors and the influence of negative transfer, so as to become familiar with Japanese pronunciation rules.

(2) Using multimedia teaching

Phonetics learning is relatively boring. In traditional teaching methods, teachers use blackboard writing, speaking, and writing as the main teaching methods. The simple forms often lead to inaccurate pronunciation and low interest in learning. Modern teaching methods such as multimedia combine text, sound, animation, graphics and other tools into one. Knowledge is transmitted through multiple sense organs such as vision and hearing, which not only guarantees the quality of pronunciation, but

also helps improve students' learning motivation and memory, so as to achieve the best teaching effect.

3 Experiment

3.1 Survey Object

This article uses two classes of Japanese majors in a university as the experimental class and the control class. Among them, there are 27 students in the experimental class, 8 boys and 19 girls; 28 in the control class, 8 boys and 20 girls. Teachers in the experimental class use voice visualization software for pronunciation teaching, while teachers in the control class use traditional Japanese classroom teaching methods. Both classes had pre-tests before the experiment, and the students' Japanese proficiency was comparable. After the 3-month study period was over, both classes were post-tested. In order to reveal students' self-perception of pronunciation through visualization technology, a questionnaire survey was conducted among students after three months.

3.2 Data Analysis

Use Microsoft Office Excel 2010 and SPSS16.0-Statistics Program for data analysis and processing. Use Microsoft Office Excel 2010 to make bar graphs to describe the data of each item in the questionnaire, use SPSS16.0-Statistics Program to calculate the mean and correlation, and use the t test to compare the effects of the two teaching methods, to verify that students are familiar with the voice visualization technology attitude of Japanese phonetics teaching. The t-test calculation formula is:

$$t = \frac{\bar{X} - \mu}{\frac{\sigma_X}{\sqrt{n-1}}} \quad (1)$$

If the sample is a large sample, it can also be written as:

$$t = \frac{\bar{X} - \mu}{\frac{\sigma_X}{\sqrt{n}}} \quad (2)$$

Here, t is the deviation statistic between the sample mean and the population mean; it is the sample mean.

4 Discussion

Table 1 shows the attitudes of students to voice visualization software in assisting Japanese voice teaching.

Table 1. Students' attitudes towards voice visualization software-assisted Japanese voice teaching

No.	Statement	Totally Disagree	Disagree	Undecided	Agree	Totally Agree
Q1	Visualization software can help me in learning pronunciation	0%	0%	10%	30%	60%
Q2	I prefer to discover prosodic rules by myself though waveform	0%	0%	20%	20%	60%
Q3	Visualization software is not difficult to use	0%	0%	20%	30%	50%
Q4	I dislike imitating the teaching or following the tape in pronunciation learning	0%	0%	0%	50%	50%
Q5	Visualization software can clearly show the prosodic features	0%	0%	10%	10%	80%
Q6	Visualization software can arouse my interest and hold my attention	0%	0%	10%	0%	90%

It can be seen from Fig. 1 that most students think that using visualization technology to assist pronunciation learning can help them distinguish the rhythm characteristics of speech; 60% of students are more willing to find the rhythm rules by themselves through the waveform in the software, and they think the visualization software can clearly show prosodic features such as pauses, stress, and intonation; most students dislike traditional Japanese phonetic teaching methods; 90% of students like the application of visualization software because it can attract students' attention and keep their attention. In summary, the application of visualization software in Japanese phonetics teaching has not only cultivated students' understanding of the phonetic system of the target phonetic, but also cultivated a positive attitude towards learning prosodic features in the classroom.

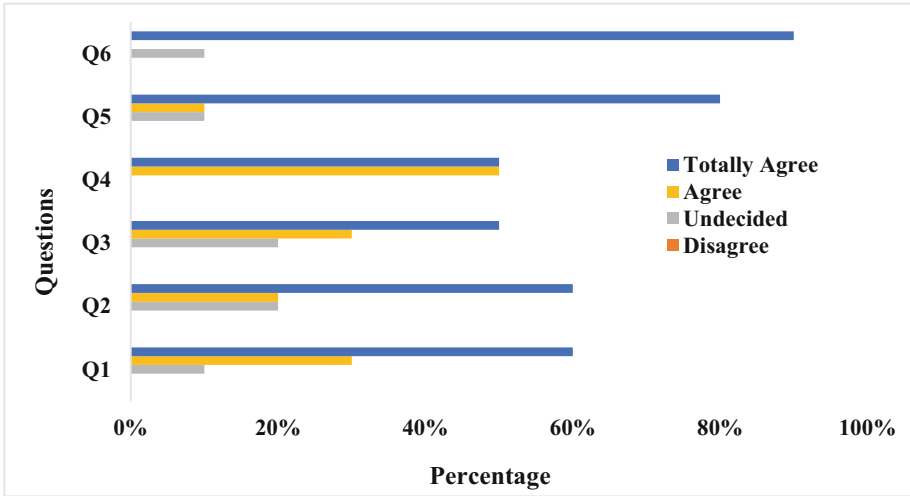


Fig. 1. Students' attitudes towards voice visualization software-assisted Japanese voice teaching

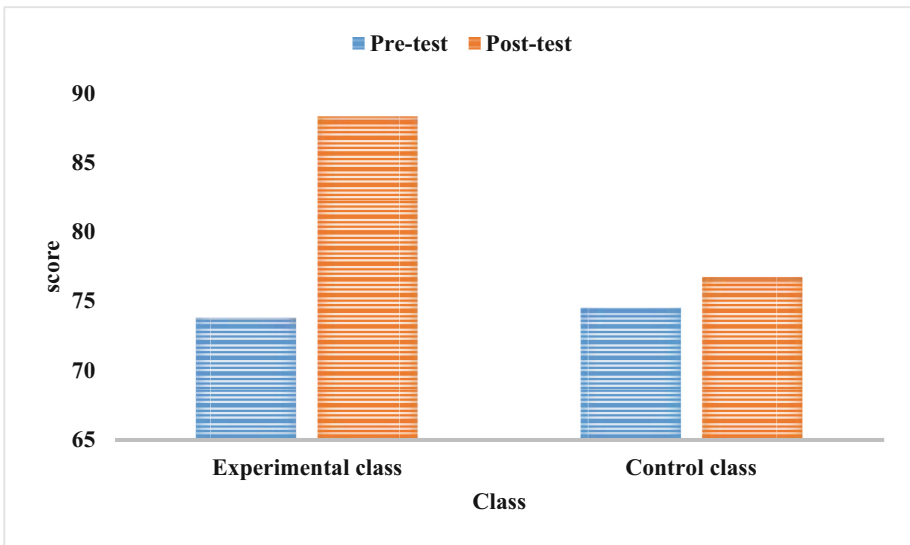


Fig. 2. Average scores of the two classes in the pre-test and post-test

It can be seen from Fig. 2 that there is no significant difference in the average Japanese voice scores of the experimental class and the control class before the experiment, but after three months of different teaching, the average score of the experimental class is significantly higher than that of the control class. Therefore, Software-assisted Japanese voice teaching based on voice visualization technology is effective and feasible.

5 Conclusions

The in-depth integration of voice visualization information technology and Japanese voice teaching must not only fully consider the rapid development of modern information technology, but also have the systematization and meticulousness of Japanese voice teaching. Only through gradual exploration and improvement, the organic integration of the two must be realized. In order to promote each other, we can better accomplish the goal of pronunciation teaching, thereby improving the teaching effect of Japanese listening and speaking in colleges and universities, thereby developing scientific college Japanese teaching.

References

1. Uihlein, M.: Maps, videos, and structures: visualizing structural concepts through media-based assignments. *Building Technol. Educator's Society* **2019**(1), 44 (2019)
2. Przedlacka, J., Ashby, M.: Comparing the received pronunciation of J. R. Firth and Daniel Jones: a sociophonetic perspective. *J. Int. Phonetic Assoc.* **49**(3), 1–20 (2018)
3. Liu, X.: A study of English phonetic teaching strategies from the perspective of embodied cognition. *Theory and Practice in Language Studies* **11**(5), 556–560 (2021)
4. Yang, W., Zhao, X.: Research on the function of visual phonetic software Praat in vocational English phonetics teaching. *J. Phys. Conf. Series* **1856**(1), 012057 (2021)
5. Karlina, Y., Rahman, A., Chowdhury, R.: Designing phonetic alphabets for Bahasa Indonesia (PABI) for the teaching of intelligible English pronunciation in Indonesia. *Indonesian J. Appl. Linguist.* **9**(3), 726–734 (2020)
6. Vronskaya, I.V., Osipova, A.S.: The issue of developing and assessing receptive phonetic sub-skills in early foreign language teaching. *Comprehensive Child Studies* **2**(2), 109–116 (2020)
7. Těpáník, S., Vlková-Mejvaldová, J.: Impact of teachers' beliefs on teaching phonetic aspects: the case of Czech as L1. *J. Lang. Cult. Educ.* **7**(2), 108–139 (2019)
8. Azarova, I., Ivanytska, I., Nykyforenko, I., et al.: Interactive online teaching of phonetic skills: introductory phonetic course. *Arab World Engl. J.* **7**(1), 94–112 (2021)
9. Perrachione, T.K., Oppenheimer, B.B.: A listening laboratory-based approach to teaching clinical phonetic transcription. *J. Acoust. Soc. Am.* **149**(4), A146–A146 (2021)
10. Meng-Wei, T.: The issue of differences in the phonetic system of Chinese in mainland China and Taiwan (in terms of teaching Chinese as a foreign language in Russia). *Prepodavatel XXI vek* (3), 173–181 (2020)
11. Ashby, P.: Does phonetics = pronunciation? 100 years of phonetics in pronunciation teaching. *ELOPE Engl. Lang. Overseas Perspect. Enq.* **17**(2), 9–26 (2020)
12. Oktavianti, I.N.: The use of phonetically reduced modals in present-day English: a corpus-based analysis. *English Lang. Teach. Educ. J.* **1**(3), 134 (2019)



Internet Financial Risk Forecast System Based on Artificial Intelligence Algorithm

Lianying Ge¹, Yachen Shen¹(✉), and K. Vijayakumar²

¹ School of Finance, Xi'an Eurasia University, Xi'an 710065, Shaanxi, China

² University of South Wales, Sydney, Australia

Abstract. Artificial intelligence is a new field that has developed in recent years. It has a certain degree of application in finance, network and biomedicine. At present, Internet finance is developing rapidly in my country. Among them, the P2P online loan platform model has the fastest and hottest development and has become a typical representative of Internet finance. This article mainly uses the experimental investigation method and statistical analysis method to study the financial and financial indicators of Internet companies, and combines artificial intelligence algorithms to model the forecasting system. Through the overall survey, it is found that the fluctuations in the various capabilities of an Internet company have fallen, and by 2021 it will fall to 0.36 solvency. This shows that with the acceleration of the Internet, financial and financial risks are becoming more and more frequent, which requires us to predict the risks. Therefore, the study of the forecasting system in this article is particularly important.

Keywords: Artificial intelligence algorithm · Internet finance · Financial risk · Forecasting system

1 Introduction

With the rapid development of Internet technology, online finance has gradually become a topic of concern. In this process, traditional financial institutions and enterprises have had different degrees of influence. On the one hand, it causes direct or indirect losses in terms of capital requirements and financing methods. On the other hand, a series of financial crises are caused by moral hazard problems caused by information asymmetry.

There are many people who study the Internet financial financial risk prediction system of artificial intelligence algorithms, and they have put forward many opinions. For example, Zhang Jingui said that the risk early warning model can more accurately measure the financial risks of manufacturing companies. It is an effective attempt to apply artificial intelligence algorithms to the field of business management and has certain practical significance for the risk analysis of financial companies [1]. Li Kehong said that artificial intelligence has brought fresh air to traditional industries and changed the way companies do traditional business. While providing beneficial help for financial management, it also provides a higher level of traditional financial concepts, financial processing systems and financial management models [2]. Ye Wentao believes that the

integration of artificial intelligence and financial risk early warning has produced an artificial intelligence model with very high prediction accuracy [3]. Therefore, the research on the Internet financial risk prediction system from the perspective of artificial intelligence algorithms has a foundation and a basis, which is profitable for enterprises.

This article first studies the relevant theoretical knowledge of Internet finance. Secondly, define and analyze the financial risk. Then the risk management is explained, and the forecasting system is designed using the BP neural network algorithm. Finally, through modeling and investigation, it is concluded that the financial risks of Internet companies are still serious, and there is an urgent need for predictive system prevention.

2 Internet Financial Risk Prediction System Based on Artificial Intelligence Algorithms

2.1 Internet Finance

Internet financing is a new financing method for the development of my country's financial industry. Most of the market investors facing the Internet financing model are ordinary investors. In Internet financing, market information can be shared by anyone. Internet finance has the characteristics of service scope. Internet funds mainly rely on big data, social media, search engines, cloud computing and other technologies to grow and grow. By using Internet software technology, Internet finance can significantly reduce the division of labor in the financial industry and the need for proprietary technology [4, 5].

The Internet is not only a secondary technical tool in financial affairs, it will first change financial transactions and existing organizational forms. The role of Internet financing is not only to stimulate the development and dissemination of financial services through the Internet. Its core is a new financing model and a new operating structure, but the main function of financing has not changed. Internet finance can use Internet technology to better serve the real economy [6, 7].

With the rapid development of Internet finance, it has seven different development models. Crowdfunding means raising money from people. Crowdfunding generally involves three parties: initiators, supporters, and platforms. Crowdfunding is when the initiator provides a creative project that lacks funds to the crowdfunding platform. The crowdfunding platform tracks and verifies the project. Once the project is approved, it can be created and invested on the crowdfunding platform to attract supporters. The essence of the P2P network lending model is to connect fund buyers and fund providers through a third-party network lending platform. The operation of P2P network lending is that borrowers and lenders can directly publish information through a third-party network platform to find people who have not used funds and want to borrow on time. Digital currency refers to the virtual currency of the network. With the development of Internet finance, a new type of virtual currency has emerged in my country: Bitcoin. This currency does not require the participation of the banking system and is a representative of virtual currency on the Internet. Big data financing refers to the use of cloud computing and other IT tools on financial service platforms to calculate unstructured big data obtained through network technology, comprehensively analyze customer information and data

of financial institutions, and accurately identify customer transactions and consumption patterns. Anticipate customer consumption trends to reduce risk. Computerization of financial institutions refers to the use of information technology to transform and reorganize existing business processes to obtain top-down comprehensive information of the entire financial institution [8, 9].

2.2 Definition of Financial Risk

Financial risk refers to the debt of an enterprise. According to its severity, it can be divided into four situations: First, the enterprise cannot repay the debt due to technological bankruptcy and accumulated interest. On the other hand, if all assets are less than a negative value, the net book value during the event may be negative; third, the company goes bankrupt, and all debts cannot be paid off after liquidation; fourth, the court filed for bankruptcy because of the inability to continue trading. For companies with financial risks, it is a bit narrow to regard bankruptcy as a default. Therefore, this article conforms to the broad definition of financial risk standards, and the production and operation of enterprises may deviate from expectations and suffer losses [10, 11].

The financing source is mainly equity financing, and the debt repayment pressure is relatively small. Therefore, it is inappropriate to carefully consider the financial risks of Internet business. This article adopts the general concept of financial risk, assuming that the financial risk of Internet business is related to various uncertain factors, including internal and external factors that lead to the failure of Internet production and operation. So it will suffer losses at some point [12].

(1) Influencing factors of Internet financial risks

1) Macro factors

Legal risk: Exploiting loopholes, breaking the law, or not clearly stipulated by the law. As my country's Internet funding is in the early stage of research, some laws and regulations have not yet been finalized. Most laws and regulations serve the traditional financial sector. Due to the rapid development of the Internet, there are few laws and regulations on Internet financing, and the legal control system is even more backward. High returns are often accompanied by high risks, which is an unchangeable law in the financial world. **Political risk:** P2P online lending has no barriers to entry, and there is no uniform industry standard to regulate behavior. Therefore, the P2P model has developed rapidly in just a few years without written policy restrictions. This proves that no monitoring and management services have allowed it to develop so fast. The P2P model is not regulated by the government, but with the development of this field, once you have your own system, there will be policies and regulations in place. **Market risk:** also known as systemic risk, it only requires advanced Internet marketing methods and concepts, and uses these advantages to quickly occupy a market position and grow rapidly. Affected

by the market environment, there is market risk. Economic risk: Internet finance is essentially the same as finance. Early warning and precise positioning of systemic risks in the business cycle are crucial.

2) Micro factors

Credit risk: The rapid development of Internet finance has accelerated the innovation of Internet financial products, and the Internet financial market is also changing. Internet finance is essentially finance. All financial products are closely related to credit investigation. Information asymmetry will appear in the financial process and the financial system is difficult to control. First of all, Internet finance is a kind of virtuality, breaking the constraints of time and space. Secondly, the reason why Internet financial products enter people's lives is only because it has the advantages of convenience, speed and high return, and people do not have a deep understanding of credit risk. **Liquidity risk:** the matching of managed asset management funds and financial product debt. **Illegal fund-raising risks:** The wealth management fund pool model is suspected of illegal fund-raising risks, unqualified borrowers' illegal fund-raising risks, and typical Ponzi scheme fund-raising risks. **Payment and settlement risk:** In a virtual environment, the payment risk of financial activities increases. **Reputational risk:** Due to inconsistent supervision, each platform uses its own standards to regulate borrowers and investors, and there are also operational errors. If venture capitalists suffer heavy losses, P2P platforms will also damage their own reputation. **Technical risks:** Cloud computing, big data and other technologies are not enough to support the huge Internet financial system. Hacking, virus penetration, financial fraud, etc. Even at this stage, this system is still full of risks. **Operational risk:** The risk of loss due to internal personnel or company program errors. **Safe operation risk:** Without a global monitoring system, risk prediction and control are still not systematic. **Information risk:** The Internet financial backend itself is virtual, and the data is very public. In addition to providing convenient services to customers, it also carries huge information risks.

(2) Risk management

Risk management is a management science that studies the rules of risk occurrence and risk control technology, and implements effective risk control and measures to properly manage losses in the management process of risk identification, assessment and treatment.

(3) Risk types of Internet finance

Compared with other industries, the financial industry has a higher risk management awareness and a more comprehensive risk management system. The Internet finance that has emerged in recent years represents a business model that is completely different from the traditional financial industry: Compared with the risks of traditional financial business models, Internet financial risks still present the following unique risk types: risk intermediate platforms, network security risks, Operational risk, liquidity risk.

2.3 Application of BP Neural Network Algorithm in Risk Assessment

BP network is a special learning network. Its working process is to first set the output value of each mode, and then input the learned storage mode from the input layer to the middle layer, and then distribute such a form of three-level forward propagation network from the middle layer to the output layer. Its structure is shown in Fig. 1.

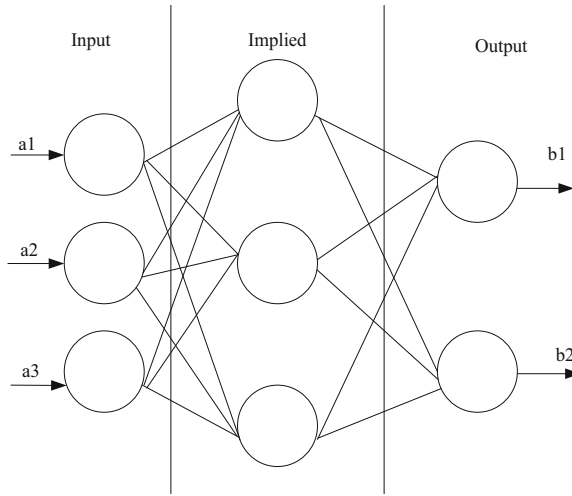


Fig. 1. Schematic Diagram of BP Neural Network Structure

When learning a BP neural network, you need to determine the basic function of the neuron first, which is generally based on a linear function, and the basic form is as follows:

$$w = \sum_{i=1}^m v_i a_i - \eta = a' - \eta \tag{1}$$

Using the BP learning algorithm, the output function can be differentiable and monotonically increasing. The connection weights of nodes usually change gradually. You can enter after correction. Usually, the sigmoid function is used to calculate BP. The form of transmitting network data is as follows:

$$g(a) = (1 + e^{-\theta a})^{-1} \tag{2}$$

After adjusting the node and determining the accurate threshold range, the error can be greatly reduced, and accurate output can be achieved, thereby efficiently completing the BP neural network operation.

3 Financial Risk Early Warning Design

3.1 Construction of Financial Risk Indicators

The analysis structure of the financial risk early warning system must be combined with the analysis of the company’s financial ratio, especially in the following four aspects.

Profitability reflects the source and level of the company's profitability, solvency is used to assess the company's debt risk status, and operability reflects the company's management level. In addition, based on the company's development capabilities, we can identify the company's future development trends and space. Before creating a financial early warning model, the review of financial indicators is also very important, which is related to the ability to effectively identify the financial risk status of the business.

3.2 Correlation Analysis of Financial Indicators

The correlation analysis of financial ratios analyzes ratios from four aspects: profitability, solvency, operability and development potential. Continue to study the four selected indicators reflecting profitability, solvency, serviceability and viability, and conduct correlation analysis.

3.3 Construction of Non-financial Indicators

The impact on company performance is gradually taken into account in many performance analyses, and there is a tendency to incorporate it into the overall evaluation system. Financial indicators cannot evaluate the company's overall performance, and only considering the impact of financial indicators will cause differences.

3.4 Construction of Financial Risk Early Warning Model

The traditional efficiency coefficient method has relatively discrete indicators for identifying financial risks, and the number of classification levels will greatly affect the effectiveness of the financial risk early warning model. The weight of financial indicators. The company's current financial status and operating performance are often reflected in financial indicators. The weighting of non-financial indicators and the financial risk warning interval can more accurately measure the company's financial risk.

Early warning process as follows. First, extract the key data from the company's annual financial statements, and then use the efficiency coefficient method to calculate the required standard values one by one. These companies are then ranked and standard index scores are assigned accordingly. Calculate the reference standard value and standard coefficient of each method according to the calculated index value and the method of the early warning system. The single coefficient of performance calculation method is obtained by the coefficient of performance method, which is calculated by combining the index value and the reference value of the gear. Differentiate the score of the aisle with the score of the top aisle, and then multiply the result by the efficiency coefficient of each index to calculate the corresponding adjustment value of LeTV. Finally, the calculated global rating factor is compared with the alert interval to identify the company's financial risks.

4 Analysis of the Application Results of the Prediction System

4.1 Evaluation of Risk Indicators for Internet Companies

This article selects the development data of a certain Internet company for analysis. Judging from the company’s 2017–2021 comprehensive evaluation coefficient results, the overall trend shows a downward trend. The details are shown in Table 1:

Table 1. Assessment of the Risk Indicators of Internet Enterprises

	2017	2018	2019	2020	2021
Profitability	0.72	0.85	0.82	0.65	0.52
Operating capacity	0.25	0.34	0.51	0.62	0.59
Solvency	0.92	0.73	0.54	0.46	0.36
Development ability	0.72	0.7	0.69	0.69	0.65
Non-financial coefficient	0.54	0.56	0.61	0.64	0.61
Comprehensive coefficient	0.65	0.62	0.59	0.6	0.5

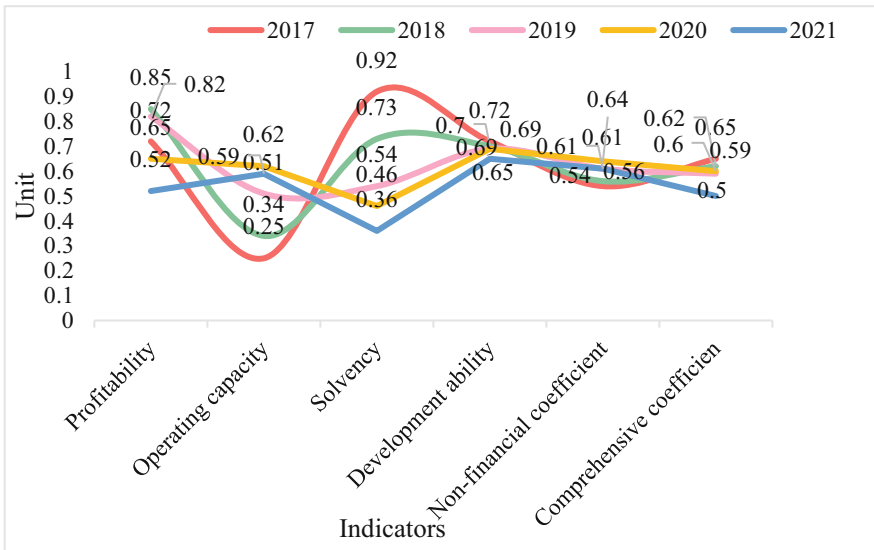


Fig. 2. Assessment of the Risk Indicators of Internet Enterprises

As shown in Fig. 2, according to the financial risk alert range, the company has been in a state of alert in the past few years. In the later period, there was a serious alert state, and it was obvious that LeTV did not effectively implement control over financial risks.

5 Conclusion

The development of artificial intelligence technology has made it a hot field, as is the financial industry. Due to the rapid rise and development of Internet finance, traditional banking services have been unable to meet customer needs and market demands. With the acceleration of network platform construction and application software development, a large amount of capital has been invested in Internet finance, which has increased the occurrence of capital risks. Moreover, according to the results of the survey, Internet companies are still in the range of frequent risk fluctuations. This strongly requires companies to have a system for risk prediction to prevent companies from failing due to insufficient risk aversion capabilities.

References

1. Jingui, Z., Fan, C., Bin, W.: Research on optimizing financial risk warning of SVM manufacturing company based on PSO. *Friends of Accounting* **000**(014), 52–56 (2017)
2. Kehong, L.: Research on smart financial management based on artificial intelligence. *New Accounting* **135**(03), 36–40 (2020)
3. Wentao, Y.: Comparison of financial risk early warning from the perspective of traditional “Z-Score” model and artificial intelligence. *Chinese Township Enterprise Accounting* **000**(011), 62–63 (2019)
4. Bordel, B., Alcarria, R., Robles, T., You, I.: A predictor-corrector algorithm based on Laurent series for biological signals in the Internet of Medical Things. *IEEE Access* **8**, 109360–109371 (2020)
5. Eramo, V., Lavacca, F.G., Catena, T., Salazar, P.J.P.: Proposal and investigation of an artificial intelligence (AI)-based cloud resource allocation algorithm in network function virtualization architectures. *Future Internet* **12**(11), 196 (2020)
6. Snoun, H., Bellakhal, G., Kanfoudi, H., Zhang, X., Chahed, J.: One-way coupling of WRF with a Gaussian dispersion model: a focused fine-scale air pollution assessment on southern Mediterranean. *Environ. Sci. Pollut. Res.* **26**(22), 22892–22906 (2019)
7. Vankevich, A., Kalinouskaya, I.: Ensuring sustainable growth based on the artificial intelligence analysis and forecast of in-demand skills. *E3S Web of Conf.* **208**, 03060 (2020)
8. Arora, M., Pal, A.: An assessment on computational intelligence based techniques in prediction of municipal solid waste management. *Test Eng. Manag.* **83**, 16471–16476 (2020)
9. Pourdaryaei, A., Mokhlis, H., Illias, H.A., Aghay Kaboli, S.H.R., Ahmad, S., Ang, S.P.: Hybrid ANN and artificial cooperative search algorithm to forecast short-term electricity price in de-regulated electricity market. *IEEE Access* **7**, 125369–125386 (2019)
10. Tkatek, S., Belmzoukia, A., Nafai, S., et al.: Putting the world back to work: An expert system using big data and artificial intelligence in combating the spread of COVID-19 and similar contagious diseases. *Work* **67**(3), 557–572 (2020)
11. Dong, M.S.: Intelligent early warning of internet financial risks based on mobile computing. *Int. J. Mob. Comput. Multimedia Commun.* **11**(2), 61–78 (2020)
12. Dolezel, P., Holik, F., Merta, J., Stursa, D.: Optimization of a depiction procedure for an artificial intelligence-based network protection system using a genetic algorithm. *Appl. Sci.* **11**(5), 12 (2021)



Structure Analysis of Digital Economy Association Network Based on Intelligent Algorithm

Jing Tan¹(✉) and Yao Gu²

¹ Department of Management, Wanjiang University of Technology, Maanshan, Anhui, China

² Department of Finance and Economics, Wanjiang University of Technology, Maanshan, Anhui, China

Abstract. In the relevant documents of the 14th Five-Year Plan, it is pointed out that China's digital development should be accelerated in order to promote the deep integration of China's digital economy and real economy. In order to deeply investigate the economic role, industrial characteristics and radiating industrial scope of digital economy industry in China, an industrial association network is constructed based on China's input-output data, and the structural characteristics of China's digital industry in the entire industrial network are analyzed by using maximum spanning tree algorithm and threshold network algorithm. It provides empirical support for the development of China's digital economy at the level of relevant economic industries. Relevant research results confirm that: first, according to the current economic development and industrial structure, digital industry has great potential and development space. In the future, it can be considered to adjust the macroeconomic structure, strengthen the promotion of digital technology to more industries, and improve its overall economic status. Secondly, in terms of the main radiation scope of digital industry, the degree of digitalization of various industries is different, which needs to be treated differently in actual policies. At present, the development of China's digital industry focuses on digital hardware manufacturing, while the development of digital software and services and digital data is insufficient.

Keywords: Digital economy · Network structure · Intelligent algorithm

1 What is the Digital Economy

In 1996, Tapscott, an American scholar, put forward the concept of digital economy, which he believed was the product of “intelligent network era”. It is not only the intelligence of network technology, but also the combination of existing economy, creativity and intelligence to create wealth and promote social development. Different scholars and organizations at home and abroad have put forward their own views on the concept of digital economy. Foreign scholars and organizations believe that digital economy focuses more on commercial activities and digital media with digital technology (Fig. 1). Chinese researchers and institutions usually build on the digital economy definition set

forth in the G20 Digital Economy Development and Cooperation Initiative. This paper focuses on the domestic digital economy related research, combined with the relevant knowledge of intelligent algorithms, while taking modern information network as an important carrier, the effective use of information and communication technology as an important driving force of efficiency improvement and economic structure optimization of a series of economic activities.

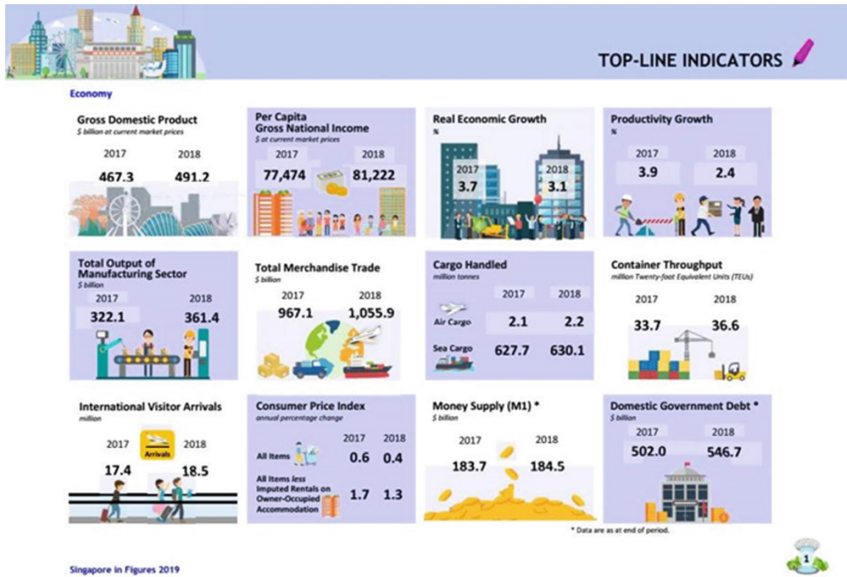


Fig. 1. Singapore’s digital economy report

In China, digital economy is divided into two types: digital industrialization in a narrow sense and industrial digitalization in a broad sense. Digital industrialization is equivalent to the traditional information industry, including the electronic and communication equipment manufacturing industry, telecommunications, radio and television and satellite transmission service industry, Internet and related service industry, software and information technology service industry in the national economy industry classification. Due to the deepening integration of information technology and other industrial sectors of the national economy, digital economic activities are generated in traditional industries, which is the part of industrial digitalization or digital economic integration.

The digital economy is growing at a high speed in all countries around the world, and its growth rate is significantly higher than that of the national economy. From 2016 to 2017, the average growth rate of the generalized digital economy (including digital industrialization and industrial digitization) in G20 countries was 8.47%, while the average growth rate of developing countries was 16.83%. Under the condition of digital economy, platform economy becomes a new production organization form different from traditional industry. A platform is an intermediary that brings together different users and an infrastructure that takes place as user activity, “a business model based on

value-creating interactions between external suppliers and customers”, or “a business model that connects two or more independent groups in a supply way”.

2 Several Common Digital Economy Industry Network Analysis Methods

Industrial network analysis method uses complex network as data structure and analysis tool to study the intricate economic relations between industries. Digital industry, as a technology-based industry type associated with many industries, its mode of action, driving industry and economic importance need to be analyzed by industrial network method.

- (1) Extraction of industrial backbone network – maximum spanning tree algorithm.
If the actual network has a large number of edges and a compact structure, it is difficult to intuitively analyze the main association relations in the network. Therefore, it is necessary to retain the strongest industrial connection edge in the network. Here, the spanning tree algorithm is used to extract the backbone network. Classical spanning tree algorithms, including prim algorithm and Kruskal algorithm, calculate the minimum connected path in the network based on undirected graph basic data. However, the actual network in this paper is directed weighted graph, that is, the relationship between the input and output of industries is oriented, which shows that the industrial production needs the input of other industries. Industrial output consumes the products of other industries. Here, the extraction goal of spanning tree is to retain the strongest structural relationship in the network, so the connection edge is selected according to the maximum weight [1–3].
- (2) Extraction method of industrial self-generating capacity.
Comparative advantage can be either from land, resources, location, population and other natural factors of production, and from the day after tomorrow can form the industrial division and form a complete set, high-quality talent, and contains the ecological system of the whole industry science and technology, knowledge, ability, while the latter is the formation of the need to continue to capital and talent into the industry field. In this sense, industrial development does not necessarily follow the path of comparative advantage, and industrial policy can play an important role in it. If a country or region enters a digital economy segment early and invests heavily, it is likely to take the lead. Guiyang has become an important big data industry gathering place in China through the early implementation of big data development strategy. Because of the uncertainty of cutting-edge technology, there is a certain risk of failure for the government to support specific industries by “selecting winners”. However, the government can, on the basis of the full analysis of experts, choose those industries that are about to enter the industrialization stage and have fewer enterprises (or regions) to support them vigorously, and guide the investment of social capital with government industrial funds and preferential policies.
- (3) Threshold network extraction algorithm.
In order to discuss the digital industry and its main radiating industry separately, it is necessary to extract and analyze the industry types closely related to the digital

industry from the original industrial network. From the point of view of actual calculation, the digital industry network nodes are taken as the center, and the threshold value is set for the connection relation around them. The nodes with strong connection relation are retained and the nodes with weak connection relation are discarded to obtain the main driving group centered on the digital industry. The calculation method is as follows: step 1: The digital industry nodes in the industrial network are selected, and the weights of the association relation between each digital industry node and other industrial nodes are arranged in order from large to small, and respectively compiled into the set of connected edges. Step 2: Set a threshold value according to the numerical distribution of the strength of all the association relation of digital industry nodes. Step 3: Build the strongly connected edge set [4].

3 Industrial Network Analysis Method of Digital Economy under Intelligent Algorithm

According to the data of China's input-output table in 2007, 2012 and 2017, the industrial networks of the three years can be constructed respectively, and the visualization results are shown in Fig. 2.

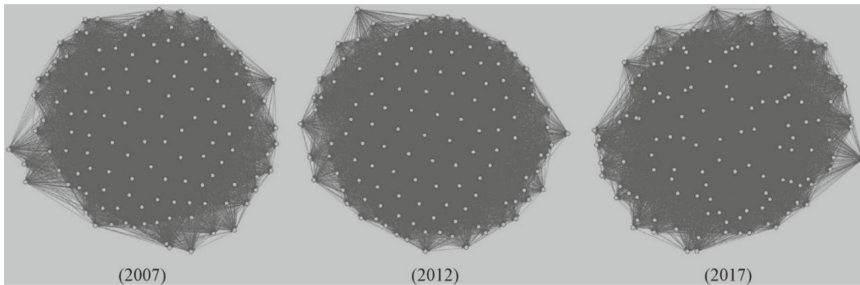


Fig. 2. Time sequence diagram of industrial network based on input-output table data

From the annual change of the number of directed edges in the input-output table, the number of associated edges in the input-output network increased from 13121 to 13404 from 2007 to 2012, indicating that the relationship between industries is more extensive and universal. In 2017, the number of associated edges of the input-output network decreased to 12,638, indicating that the input-output relationship between industries is weakened, production relations are more intensive and endogenous, intermediate inputs are more “pure” than before, and production links are clearer and more specialized [5].

In view of the connotation and development trend of digital economy, scientific measurement and evaluation of the development level of digital economy is not only directly related to how to objectively measure the impact of digital economy on the overall economy, but also will further affect the overall judgment of the macro situation and the accurate formulation of relevant policies. This paper follows the existing research and adopts the digital economy development index in the Digital China Index Report

2019 released by Tencent Research Institute. The release of the index provides key materials for the study of the development status and impact of China's digital economy, and scientifically reflects the development level of digital economy in various regions.

At present, the digital wave is sweeping the world, and China's digital technology is also accelerating innovation, constantly promoting the rapid emergence and rapid growth of new business forms and new models, making digital economy with new development trend and changing characteristics (Table 1). As represented by 5 g of a new generation of high-speed information communication technology rapid development and wide application, around the acceleration of digital new infrastructure facilities construction, high speed and mobile, security, perfecting ubiquitous digital key infrastructure, effectively supporting the Internet, big data, artificial intelligence and the depth of the traditional economic integration, Thus rammed the digital economy development infrastructure. Digital application has greatly extended and expanded the boundaries and fields of traditional industries. Digital industry, digital life, digital government affairs, digital culture and so on have shown a trend of accelerated development, which not only optimizes the level of organization and management, but also greatly improves the efficiency of resource allocation and social governance rate [1, 6]. Especially digital industrialization and industry in-depth reform of the digital, big data and cloud computing, Internet of things, such as artificial intelligence digital technology to traditional industry's active role is increasingly prominent, an increasingly important role in the configuration of the data elements, new forms, new industries and new patterns emerge and become the leading new momentum of economic development, the sustainable development of the economy growing (Table 2).

Table 1. Digital economy index meaning

Index	Meaning
Total digital Economy index	To demonstrate the level of development of the digital economy as a result of the deep integration of the real economy and all sectors of society
Digital Industry Sub-index	The synthesis of all digital cultures presented from the consumer side
Digital Culture sub-index	Show the ability and level of consumption culture throughout the country
Digital government sub-index	Measure the level of government affairs in all parts of the country

Disruptive innovation constantly emerging, platform of economy and the speeding growth, network effects and the "winner-take-all", "dandelion effect" with new features such as ecological competition make active and fast-growing digital economy innovation, has become one of the important driving force of old and new kinetic energy conversion, especially for the backwardness of countries and regions are seeking new kinetic energy, speed up the economic development opportunities. Further promote the development of digital economy and accelerate the transformation of old and new drivers of growth [7].

Table 2. China's digital economy in 2020

	The mean area	The maximum	The minimum value	The standard deviation
The national	20.236	105.710	1.770	18.902
In the east	32.860	105.710	6.080	25.497
In the middle	16.915	29.930	8.73	6.941
In the west	10.879	29.750	1.770	7.348

4 Analysis of Radiative Effect of Digital Economy Industry Network

Digital software and service industries (telecommunications and other information transmission services, software and information technology services) Network node strong connection Non-digital industries are mainly service industries, including construction, wholesale and retail, finance, insurance, real estate, education, health, public administration and social organizations. Therefore, the digital software and service industries are mainly driven by the service industry, which has little direct impact on the manufacturing industry. The digital data industry consists of two subsectors, in which the news and publishing industries are connected to the digital industry threshold network through telecommunications and other information transmission services, and radio, television, film and audiovisual recording operations are separate from the other digital industry main networks. The strongly related industries of digital data industry are also service industries, including construction, wholesale and retail, finance, real estate, research and experimental development, education, health, public administration and social organizations [6]. In order to more clearly and quantitatively describe the main correlation between digital industry and non-digital industry, the main correlation industries and correlation weights of digital industry are presented in table form here. As there are many types of digital industry subdivisions, this paper selects an industry category from digital hardware manufacturing, digital software and services, and digital materials as representatives for analysis, and the selected industries are computer, software and information technology services, news, and publishing. The threshold network constructed based on this idea can clearly reflect the connection relationship between digital industrial nodes, and radiate from digital industrial nodes to the main association relationship of non-digital industrial nodes in the next layer [8, 9].

5 Conclusions

The comprehensive development of digital economy based on intelligent algorithm not only focuses on the purchase and upgrade of digital hardware in ordinary industries, but also needs to strengthen the application of software customization, information technology and Internet technology. At the same time, it takes into account the knowledgeization of massive data in economic activities to achieve a stronger radiation and driving role of digital industry. From the perspective of economic network, China's digital industry

plays an important role in the whole economy. The network association structure of digital economy based on intelligent algorithm can effectively formulate the development of China's digital economy and provide the basis for decision-making of relevant policies.

Acknowledgements. This work was financially supported by Key discipline construction project of Wanjiang University of Technology in 2019: International Trade(WG XK19002), University scientific research project of Anhui Provincial Department of education in 2020: Study on the influence of the "One Belt, One Road" trade facilitation level on Anhui's trade flow (SK2020A064).

References

1. Paliszkiwicz, J., Chen, K.: Trust. In: Organizations and the Digital Economy: Theory and Practice. Taylor and Francis (2021)
2. Almunawar, M.N., Islam, M.Z.: Patricia ordóñez de pablos. digitalisation and organization design. In: Knowledge Management in the Asian Digital Economy. Taylor and Francis (2021)
3. Almunawar, M.N., Islam, M.N.: Patricia ordóñez de pablos. digital transformation management. In: Challenges and Futures in the Asian Digital Economy. Taylor and Francis (2021)
4. Yan, Y., Wang, B., Zou, J.: Blockchain. In: Empowering Digital Economy. World Scientific Publishing Company (2021)
5. Zani, B.: Women migrants in southern china and taiwan: mobilities. In: Digital Economies and Emotions. Taylor and Francis (2021)
6. Popoff, A.: Fundamentals of signal processing in generalized metric spaces. In: Algorithms and Applications. CRC Press (2021)
7. Almunawar, M.N., Islam, M.Z.: Patricia ordóñez de pablos. digitalisation and organization design. In: Knowledge Management in the Asian Digital Economy. Taylor and Francis (2021)
8. Almunawar, M.N., Islam, M.Z.: Patricia ordóñez de pablos. digital transformation management. In: Challenges and Futures in the Asian Digital Economy. Taylor and Francis (2021)
9. Banerjee, A.: Flexible Multibody Dynamics: Algorithms Based on Kane's Method. CRC Press (2021)



The Digital New Infrastructure Construction System of Provincial Power Grid

Xiao Lu¹, Lei Wei¹, and Zhe Wang²(✉)

¹ State Grid Jiangsu Electric Power Co., Ltd., Nanjing City, Jiangsu Province, China

² State Grid Jiangsu Electric Power Co., Ltd., Economic and Technological Research Institute, Nanjing City, Jiangsu Province, China

Abstract. “New infrastructure” is the strategic basis for the development of the digital economy. It is based on the construction of digital infrastructure. 5G, artificial intelligence, and big data are commonly known as “digital new infrastructure”. This paper studies the digital new infrastructure construction system of the provincial power grid, and understands the relevant theories of the digital new infrastructure, then constructs the data security system construction of the provincial power grid digital new infrastructure, and finally detects the data retrieval in the data security construction. According to the detection results, the original traversal search method can also be used to quickly find the target data. However, with the gradual increase in the number of users, the advantages of this solution are becoming more and more obvious. Searching the catalog can directly find the corresponding data block to be accessed.

Keywords: Smart grid · Digital infrastructure · Digital technology · Infrastructure construction

1 Introduction

The work report of the State Council clearly stated that the construction of new infrastructure should be accelerated [1, 2]. Among them, 5G, artificial intelligence, big data and industrial Internet are called “new digital infrastructure” [3, 4], which means that the latest intelligent digital information technology is used to create the bag industry pattern in the 4.0 era and create an advanced infrastructure system [5, 6]. Digital new infrastructure occupies half of the seven core areas of new infrastructure. Its importance is self-evident, but the actual progress and implementation process are not smooth. The key is the lack of preparation and necessary motivation [7, 8].

Regarding the research on power grid infrastructure construction, some researchers have pointed out that there are many problems in the management of existing infrastructure. One of the best ways to solve these problems is to establish a complete power grid infrastructure management and control system. Managing grid infrastructure projects is very complicated. During the project, there is a large amount of data that needs to be processed quickly, and the quality, cost, schedule and other implementation results of the project must be controlled to provide a basis for subsequent steps and future manager

decisions. Only advanced computer and information technology can fully accomplish this task. The implementation of grid infrastructure management is the most important feature of management software, and it is also a sign of automation, modernization, and internationalization of management methods. The use of grid infrastructure management and control system can improve the intelligent level of infrastructure management of grid companies [9]. There are also researchers who conduct in-depth demand surveys around the basic business functions of integrated infrastructure project management based on the business process of the integrated infrastructure system. Based on the demonstration of the system design goals, through the combination of B/S and C/S structure and the use of integrated PowerBuilder, security has been carefully considered in terms of performance, stability and scalability, and an infrastructure based on unique business processes has been designed and implemented Information management system [10]. In addition, some scholars said that the new generation of information infrastructure runs through the entire process of data collection, decision-making, and implementation. The data center stores, calculates, and processes all data in the process of technology and business implementation. The trend of global business “digital transformation” to accelerate the transformation to the cloud is becoming more and more obvious. Cloud security has become a security battlefield, and cloud migration is the “best solution” for solving security problems in the digital age [11]. To sum up, the research on the construction of new digital infrastructure has attracted much attention, and most of them have concentrated on the research on how to implement the new digital infrastructure, and there is less research on the construction of the construction system.

This paper studies the digital new infrastructure construction system of the provincial power grid, analyzes the key technologies for the promotion of digital new infrastructure and the construction measures of the provincial power grid digital new infrastructure based on the literature, and then analyzes the data security system of the provincial power grid digital new infrastructure, and finally use experiments to verify the feasibility of the construction plan.

2 Research on Digital New Infrastructure

2.1 Key Technologies for Advancing Digital New Infrastructure

(1) Information integration

High-level information is one of the key features of digital power networks [12]. The information integration in the digital power network focuses on the horizontal integration and vertical connection of information, so the effective integration of information occurs in the entire network.

Sharing data resources is the foundation for power energy companies to achieve application integration. Constructing a comprehensive data integration application platform can not only integrate and optimize the information resources of the original application system, but also develop advanced analysis and decision-making applications to meet current and future challenges. The goal of information integration is to understand, analyze and apply all aspects of information in a timely and accurate manner to improve

the timeliness and accuracy of analysis, judgment and decision-making, and to manage and control the entire process of operation.

(2) Application integration

The digital grid uses an information integration platform to integrate and exchange information. The grid company also has a number of enterprise resource management systems, including power generation management, marketing services, engineering management, human resources, finance, and material management. In addition to exchanging information, these systems also need to be integrated into business processes, the so-called cross-border integration of business applications.

(3) Information security

Digital power network is a technological system that integrates advanced information. Substation automation systems and dispatch automation systems with real-time and online capabilities are essentially distributed information systems. The most important thing is that there are vertical information connections and interconnections between the upper and lower dispatch centers, and real-time monitoring systems such as dispatch automation are horizontally integrated with production control systems and marketing services. Information security threats under the background of database power networks are increasing. On the one hand, it must meet the security control and protection requirements of the national secondary system, and on the other hand, it must meet the information security requirements. Therefore, new information security protection technical solutions and measures must be designed according to the architecture of the digital grid.

2.2 Provincial Power Grid Digital New Infrastructure Construction Measures

This article analyzes the information security in the construction of the new digital infrastructure of the provincial power grid, and puts forward the following suggestions:

In the era of the Internet of Everything, the flow and use of data is the driving force of social evolution. Digital trust was born under the general trend of “cloud security services, data encryption chains, and service coexistence”. From peer-to-peer trust to national trust, to social media agency contract trust, the traditional trust in experience, law, ethics, and government mechanisms is being transformed into today’s code encryption and algorithms. The technical approach to this transformation is zero trust. The “Zero Trust” strategy is to not trust anyone unless the network clearly knows the identity of the person who accepts that person. There are no privileged users, traffic, systems, and sites in the “Zero Trust” strategy, so that it can avoid being based on accounts, identities, storage, etc., create personal digital rights archiving relationships. Regardless of the IP address change method or role (access subject change, etc.), the system adopts a robust security verification method to verify the release of trust at the level of ensuring that the user is always safe (verified). Within the scope of security control, a similar real world will eventually be realized: a universal method based on single face authentication, creating a trust transition from the real world to the virtual world, and seamlessly connecting online and offline “digital trust”.

2.3 Data Security Algorithms in the Construction of New Infrastructure

The secret homomorphism technique is based on algebraic theory. The basic idea is as follows, assuming that E and D represent encryption and decryption functions, the elements of the plaintext data space are a finite set $M = \{M_1, M_2, \dots, M_n\}$, and a and b represent numerical operations). Once (1) is established and (2) is established, the family of functions (E, D, a, B) is called a secret homomorphism.

$$\alpha(E(M_1), E(M_2), \dots, E(M_n)) = E(\beta(M_1, \dots, M_n)) \quad (1)$$

$$D(\alpha(E(M_1), E(M_2), \dots, E(M_n))) = \beta(M_1, \dots, M_n) \quad (2)$$

3 The Construction of the Data Security System for the New Digital Infrastructure of the Provincial Power Grid

3.1 Analysis of Data Security System Construction

In the digital age under the “new infrastructure”, smart power terminal data has become a new means of production, creating new industrial and social values, and affecting the development of the entire era. Data analysis of terminal equipment not only allows the company to adapt to the ever-changing trend of the times, but also predicts future growth trends and gives it a competitive advantage. The prerequisite for realizing the value of smart power terminal equipment data is to correctly understand, manage and use the life cycle of these data. It is used to manage the data flow and coverage of the power information system and the entire process from installation to initial storage to subsequent dismantling. Therefore, value mining must be conducted on key issues such as data storage integration, data sharing, and data protection of smart power terminal equipment, as shown in Fig. 1.

3.2 Data Storage Module

In this article, the user is a safe user identified by the power service provider, the private cloud operating environment of the power system is reliable, and the research mainly protects user-generated data in public places. Therefore, based on the combination of the smart grid hybrid cloud structure, this paper proposes a smart grid data security storage model under the hybrid cloud development mode.

In this storage model, after the user completes the required storage data encryption process, the data is transferred to a third-party cloud service provider for storage. The key used by the user to encrypt the data is the public key passed locally, and the information provided by the energy service provider after being encrypted, it will be sent to the local energy service provider for storage. As a trusted third party for users, local power suppliers can provide users with primary storage services and regular data integrity verification services. The third cloud service provider is mainly responsible for storing user data, receiving integrity verification requests issued by the local power supplier, and sending the verification result to the local power supplier.

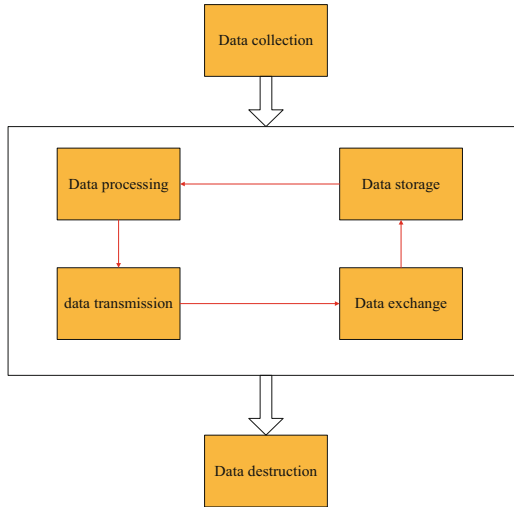


Fig. 1. Data flow of intelligent power terminal equipment

3.3 Data Sharing Module

There is a two-way flow of information between grid users and the company. In other words, access and share information. The smart grid has smart meters and various sensors. According to the storage method described in this article, the collected network data is stored on the cloud server through the data collection database station of the alliance chain. Electric power companies or distributed power users need to access cloud server data for analysis and judgment in order to formulate electricity prices and electricity usage strategies. In the traditional data sharing method, in a series of electricity transactions on the smart grid, distributed power sources use public keys to encrypt the available electricity and electricity prices. When the encrypted text is sent to the cargo user, the encryption is performed, and the user receives and completes the decryption. Decrypt the previously used private key and retrieve the plaintext shared data. This method has great limitations. If many users want to access their data, the data owner will have to perform multiple encryption operations, and the sharing process is still “one-to-one”. To use this format, you need to take into account that the ciphertext transmission sequence is cumbersome, the amount of calculation is large, and the performance is poor.

This paper designs a data access process based on intelligent data storage solutions. The intelligent network data sharing process is as follows:

Access control strategy: This is the core access control strategy of the shared method to ensure access security by verifying the identity of the user. For example, for policy = < Bill, power, check >, Bill is the owner of the power. Power means access to power files, and check means user-defined access points (such as account-defined transactions) to check power, electricity prices and other information. The program’s access control strategy includes data holders and access goals that can be achieved by writing smart contracts.

During the data access process, the visitor needs to verify the access control strategy, and retrieve the key and index values after passing the verification.

It can be seen from the data storage process that the transaction information in the block includes the information index number, the added address value, and the hash value fragmentation.

Grid data index number: Each index number can be searched through a single index, and the index number is not repeated, when users perform data queries.

The address value of the additional grid data: The index and the address value stored in the blockchain have a “one-to-one correspondence”. Users can search for the location of encrypted data stored on the cloud server based on the address value.

Grid data fragmentation value: The ciphertext is calculated by the data storage process and stored in the private cloud. After requesting the ciphertext of the data according to the address value, the ciphertext hash value of the query is calculated, and the integrity of the ciphertext data can be verified by comparing the two hash values.

3.4 Data Protection

This solution can ensure data security, save files in encrypted text form, and encrypt data before uploading, so as not to disclose the most sensitive data. This can be achieved by increasing the security of the master key. The main reason is that the encryption key is created by the master key and the block access control policy cannot be changed at will. The legitimacy of the requesting shared user can be achieved by verifying the access policy. Therefore, if the access control policy is not met, the attacker will not receive the decryption key from the data owner or the accounting node, even if it receives the usage log information for encrypted data, the accounting node can also verify the authentication certificate issued by the system through the signature of the certificate. Therefore, even if it is a forgery attack, there is no information available. The encryption algorithm used in this solution is a symmetric encryption algorithm. Each status monitoring data on the smart grid corresponds to a key to further protect the data, and the data can be decrypted to complete the transmission. Therefore, the key plays a very important role. In this article, the master key can generate sub-keys, and the master key is stored by the user.

4 Algorithm Detection

The design uses a directory search method to improve the efficiency of data retrieval. If you want to obtain specific data, you can directly find the corresponding data by searching the catalog item. However, traditional search methods usually need to traverse the data in the block until useful data is found. Compared with the traditional data retrieval method, the directory search method carries an extra amount of data, but it greatly improves the search efficiency. This article introduces traditional retrieval schemes for comparison. Facts have proved that this program greatly improves the efficiency of data recovery. Next, in order to verify the superiority of the scheme, through computer simulation experiments, the simulation results are shown in Table 1.

Table 1. Computer simulation experiment simulation results

	Our scheme	Other scheme
5	150	150
10	160	170
15	155	200
20	200	220
30	220	240
40	400	410
50	410	600
100	900	1300

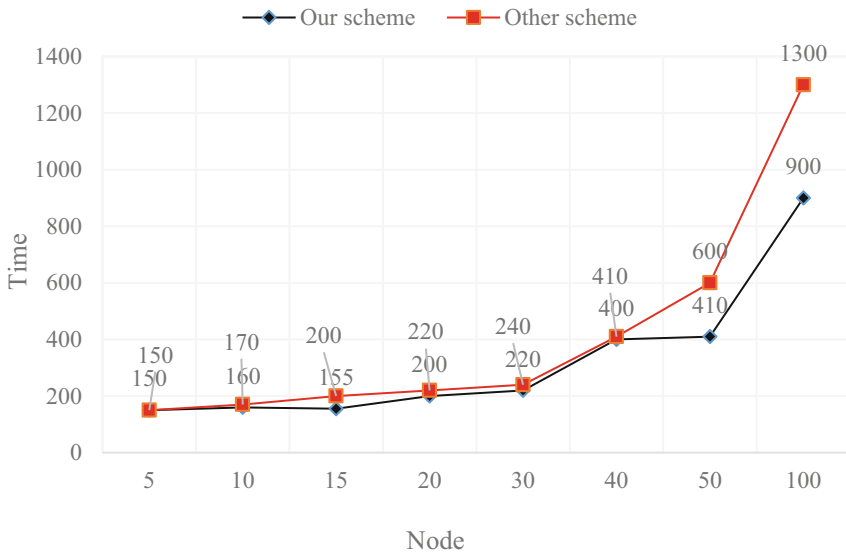


Fig. 2. Computer simulation experiment simulation results

It can be seen from Fig. 2 that if the number of nodes in the system is small, the effective data of each node accounts for a higher proportion of all data. The target data can also be quickly found using the original search method. However, as the number of users grows, the advantages of the solution described in this article become more obvious, it directly searches the catalog to find the corresponding accessible data block. For each node, the efficiency cannot be limited, even if its effective data accounts for a very low percentage of all data.

5 Conclusions

This paper studies the digital new infrastructure construction system of the provincial power grid. After understanding the relevant theories, the data security system in the digital new infrastructure of the provincial power grid is designed, and the data retrieval in the security system is tested. The test results are Searching the catalog can directly find the corresponding data block to be accessed. Even if for each node, its effective data accounts for a low proportion of all data, it cannot limit the efficiency.

References

1. Shirowzhan, S., Tan, W., Sepasgozar, S.: Digital twin and cybergis for improving connectivity and measuring the impact of infrastructure construction planning in smart cities. *Int. J. Geo-Inf.* **9**(4), 240 (2020)
2. Liu, C., Yang, J., Qiang, R., Yuan, L., Mahato, N.K.: New infrastructure-oriented power network security supervision innovative mechanism and system architecture. In: 6th International Conference on Advances in Energy Resources and Environment Engineering (2020)
3. Wang, & Jun.: New infrastructure, new engine. *Beijing Rev.* **63**(21), 38–39 (2020)
4. Blair, S.: Digital Twin Support and Open Source Tools Promised. *Engineering news-record* **281**(12), 53–53 (2018)
5. Foka, A., Misharina, A., Arvidsson, V., et al.: Beyond humanities qua digital: spatial and material development for digital research infrastructures in HumlabX1. *Lit. Ling. Comput.* **33**(2), 264–278 (2018)
6. Revell, T.: Create a digital Geneva Convention. *New sci.* **234**(3127), 21–22 (2017)
7. Tyler, N.: The advent of the digital railway. *New Electron.* **49**(12), 18–19 (2016)
8. Lobo, S., Whyte, J.: Aligning and reconciling: building project capabilities for digital delivery. *Res. Policy* **46**(1), 93–107 (2017)
9. Chamlou, N., Toczaer, C.: Entebbe makes strides in cargo infrastructure, digital processes. *Air Cargo World*, **108**(11), 12–12 (2019)
10. Vonderau, P.: The spotify effect: digital distribution and financial growth. *Telev. New Media* **20**(1), 3–19 (2019)
11. Tuamsuk, K., Subramaniam, M.: The current state and influential factors in the development of digital literacy in Thailand's higher education. *New Libr. World* **118**(5–6), 235–251 (2017)
12. Prakasha, K., Muniyal, B., Acharya, V., et al.: Efficient digital certificate verification in wireless public key infrastructure using enhanced certificate revocation list. *Inf. Syst. Secur.* **27**(4–6), 214–229 (2018)



Application of Digital New Infrastructure Business Scenarios in Power Grid Enterprises

Qinwei Dong, Chengling Jiang, and Zhoujun Ma^(✉)

State Grid Jiangsu Electric Power Co., Ltd., Nanjing City, Jiangsu Province, China

Abstract. New technologies and new business formats driven by new infrastructure will drive the deep integration of digital technology with traditional power grids, deep integration of the upstream and downstream of the energy industry chain, and accelerate the creation of first-class enterprises, creating new opportunities. This article aims to study the application of digital new infrastructure business scenarios for power grid enterprises. Based on the analysis of the application direction of infrastructure business and the information content of the new infrastructure of the power grid, combined with the analysis of the non-functional requirements of the system, the new infrastructure information management system is designed and implemented. Finally, the performance of the system was tested. The test results show that the system runs well and meets the requirements.

Keywords: Power grid enterprises · New infrastructure · Business scenarios · Information management

1 Introduction

New infrastructure is the construction of new infrastructure. New infrastructure, as the modern thought and concept of our country's economic and social development, requires scientific and technological innovation activities as the first driving force for economic and social development, and the Internet is the possibility of achieving high-quality economic and social development [1, 2]. Based on the development of information networks and the digital research of information technology, a comprehensive infrastructure and product system with intelligent integration and innovative service capabilities [3, 4].

Grid infrastructure project management is not only the highlight of the daily operation of power enterprises, but also an important source of corporate capital construction. This not only reflects the effective management of enterprises to a large extent, but also relates to the smooth progress of asset management life cycle strategic goals. At present, the scale of power grid infrastructure construction is huge, and the management of power grid infrastructure is facing huge challenges. To this end, under the new situation, actively reform the power grid infrastructure management system, enhance the awareness of the necessity of power grid infrastructure management, carefully analyze the problems that arise in the power grid infrastructure management, and promote the construction of power grid projects. Power grid infrastructure management standards and improved infrastructure management methods [5, 6]. In the 1970s, foreign

countries began to develop project management information systems. By the 1980s, project management information increased rapidly. Since then, with the development and widespread use of computers, the project management information system has continued to develop and improve [7, 8]. In the mid to late 1980s, the engineering project management system developed from single project management to multi-project management, from DOS-based character software to complete graphics software. At that time, many engineering project management systems only considered individual needs in the project management process [9, 10]. At that time, there was no software company of a decent scale, and the software and hardware platforms and development methods were relatively late, so the software developed at that time was not actually executed very well. Until the late 1980s, most software vendors focused on functional integration. Beginning in the 1990s, the development of computer software and hardware gradually matured, and many software companies were established and gradually expanded. Project management software developers began to integrate more functions. In the mid-1990s, the Internet-based project management system appeared abroad and has been continuously recognized by all parties in project construction [11, 12]. Project management software for integration, personalization, intelligence and networking is currently being developed. Most software is open, supports back-end databases, and provides better software and system inheritance.

On the basis of consulting a large number of relevant references, this paper designs a new infrastructure information management system for power grid enterprises based on the direction of infrastructure business application, the informatization content of the new infrastructure of the power grid, and the non-functional requirements of the system. The system mainly includes four modules. They are project management function module, cost management function module, schedule management function module and technology management function module. After implementing the above four modules one by one, this article tests the performance of the system to verify whether the system meets the requirements of this article.

2 Application of Digital New Infrastructure Business Scenarios in Power Grid Enterprises

2.1 Application Direction of Infrastructure Business

- (1) Deepen business applications based on the current status quo

The construction direction of this project is suitable for power grid companies that have formulated preliminary industry construction standards and have relatively mature industry applications. Its industry applications are not only well adapted to current industry needs, but also to future industry needs. Business applications that meet the above conditions are often business applications that have invested good experience in the early stages of the project. When the company is accustomed to carrying out construction based on the status quo, it can also fully protect the investment in the project and the learning experience.

(2) Transformation on the basis of existing business

The business application system of this construction direction, the current industry system construction situation is relatively unsound, the overall construction direction of future projects is relatively uncertain, and the industry application system that is seriously unbalanced with the current construction situation and the level of the grid company is more suitable for industry use. Business applications that meet the above conditions are generally business applications that have different levels of investment, or use business applications that build new transformation directions on existing infrastructure.

(3) New construction of the project

This construction direction is suitable for business applications that power grid companies have not yet set foot in or are under construction, that is, management needs. Business applications that meet these requirements typically make little or no investment in existing investments. Therefore, the use of new construction directions will not waste commercial investment.

Therefore, each power supply company must reasonably schedule the existing business processes of the industry and the basic system according to the current operating conditions and the actual conditions of the enterprise, and reorganize and optimize the basic processes of each industry in order to further improve the effect of approval tasks. Effectively promote the working process of different process links in a node in the system service. In response to the current demand for information infrastructure services, the information infrastructure service system is adjusted and improved to improve the coverage rate of the information infrastructure service system in actual work, so that the information-based auxiliary management work of the power grid infrastructure project management can be obtained.

2.2 The Informatization Content of the New Infrastructure of the Power Grid

(1) Informatization of document and data management

Taking full account of the complete life cycle of the project, a database system was established to save the entire project information, and automatically save the information according to the time node, so as to facilitate the later collection of information. Establish a database of project-related information to facilitate sharing and utilization.

(2) Engineering information processing and visualization

The amount of information and data of infrastructure projects is huge, and the relationship between various types of data is intricate. From the rich data stream, to obtain a large amount of useful data for project management, and to analyze the correlation between various data types, it is essential to make accurate decision-making for project management. Through visual information such as project management overview, project

progress, technical obstacles, safety and quality issues, project managers can view the content of the meeting and participants, so as to grasp and manage the relevant situation of the project, thereby improving efficiency.

(3) Establish effective communication channels through informatization

Establish an Internet-based information management system to allow project participants to send and receive information through the Internet and electronic media, allow project participants to communicate remotely, establish a communication and collaboration platform, and solve problems in a timely manner.

(4) Informatization of process control

Establish a real-time monitoring system for remote consulting experts and remote management project sites. The project management process, the establishment of a scientific pre-inspection system, tracking the processing of project information, predicting possible problems in the next step, and the parties involved in the construction coordinate plans and make decisions to avoid major losses.

2.3 System Non-functional Requirements

(1) Interface requirements

The style of the interface should be friendly, beautiful and consistent. It provides users with different help and instant information in different ways. The prompt information should be concise and clear, and the style of the message box should be neat and consistent. The layout and layout of functions in the navigation bar, menu bar, status bar, and home page must be logical. All homepage frames provide floating positions, and each space position on the host can be changed at will. The style and layout of the pop-up page should be designed uniformly, and the size of the pop-up page should be adjusted according to different business characteristics. Under normal circumstances, do not replace the help-related information under the normal operating conditions of the pop-up page.

(2) Safety requirements

Security requirements generally refer to the security requirements of the operating system itself, regardless of the following physics, network systems, hosts, anti-virus software, firewalls, etc. The product provides a linked CA digital certificate, which controls security according to the product principle and is not easy to invade. At the same time, the product also provides a functional log management function to protect the user's data process, even if there is an error and will not cause data loss.

(3) Functional operation response to demand

For system operation, according to the current unified requirements of enterprise business management software, the server operating system response time of this system is 3 s, including system operation report requests, logic processing and server feedback, system operating system upgrades, etc., which must be less than 3 s to satisfy the enterprise Internal software performance standards.

3 Experiment

3.1 Project Management Function Module

(1) Project management process

Project management steps: system users log in to the system through system login authentication, and the system loads the corresponding system permissions according to the user ID. After loading the system permissions, the system user clicks on the project management module, and the system loads the corresponding Project.jsp interface. After the system enters the project information, the system verifies whether the information entered by the user is valid. When the system user clicks the Save button, the system calls the business logic processing class and loads the corresponding SAVE method to save the information entered by the user. When the database operation is completed, the post-operation flag is the system that completes the project management business process.

(2) Project management function realization

Project management is the data entry of the integrated infrastructure information management system, and it is also a basic function of the system. The main function of project management is to enable infrastructure projects to be constructed in accordance with project quality standards and contract requirements. In the process of project management, the infrastructure construction unit shall carry out construction in accordance with the requirements of the contract, in accordance with the quality, procedures and schedule of the project, monitor the progress of the infrastructure project in real time, and complete the project in accordance with the regulations. The project management mode interface contains three function buttons: add, edit and query. If system users need to edit project information, they can enter the corresponding project name, start time, result time, status and other query information. After completion, the project information can be changed.

3.2 Cost Management Function Module

The main functions of the cost module include: feasibility study evaluation management, progress payment and change management, contract price management and other functions. After the user logs in to the cost management interface, he selects the new construction cost standard database, and the department head conducts mechanical analysis and compares the cost information with the previous data in the database. The operation completes the overall business cost management operation process. After a series of operations are completed, the system cost management report is generated, and finally returns to the main interface of the system.

3.3 Progress Management Function Module

The implementation of the schedule management function includes the basic schedule management class `PlanManagement`, the schedule management business logic class `PlanLogic`, and the schedule analysis implementation class `PlanAnalyse`. The schedule management base class `PlanManagement` defines the public methods of schedule management, such as `GetPlan()`, query schedule method `queryPlan()`, etc. The schedule management business logic class `PlanLogic` and the schedule analysis implementation class `PlanAnalyse` are both the basis of schedule management. Class and define each specific business processing method, such as a scheduling business. Logic class execution scheduling method `executeplan()`, scheduling save method `Executesave()`, etc. When the system user initiates a business request, the system obtains scheduling information through the `GetPlan()` method, and calls the `executesave()` method to save the scheduling plan.

3.4 Technical Management Function Module

The technical management section is divided into technical standards, technical management, and technical management. The three main aspects summarize the data of standard differences, construction plans, technical notifications, change implementation inspections, and technical project plans. Users can search for relevant data in the form of charts and tables, and provide management and approval file viewing functions. The sub-module division diagram of the technology management module is shown in Fig. 1:

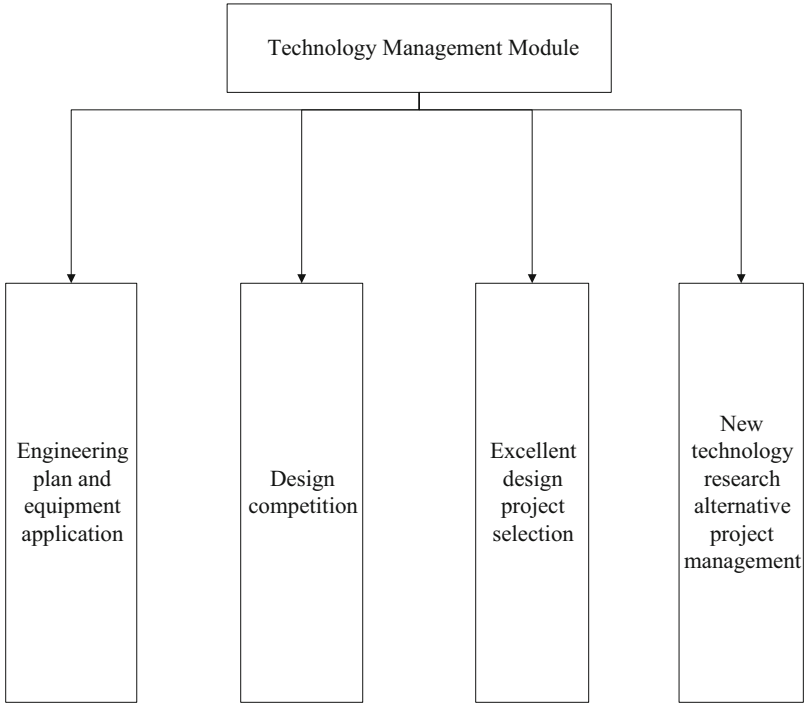


Fig. 1. Technical management module diagram

4 Discussion

In addition to functions, it also pays attention to whether the performance meets the design requirements to evaluate whether the system is stable and reliable. Performance testing is a necessary step of system testing, the purpose is to meet the needs of design and users. During the test, a client computer simulated by a “virtual” user and multiple real clients were collected to verify the operating status of the system under heavy load. Determine the critical load on the system and check whether the system performance can meet the design requirements. LoadRunner is a widely used system testing tool. Simulate system user operations, test system performance, and find system operation problems. The calculation of the average number of concurrent users and the peak number of concurrent users is shown in formulas (1), (2):

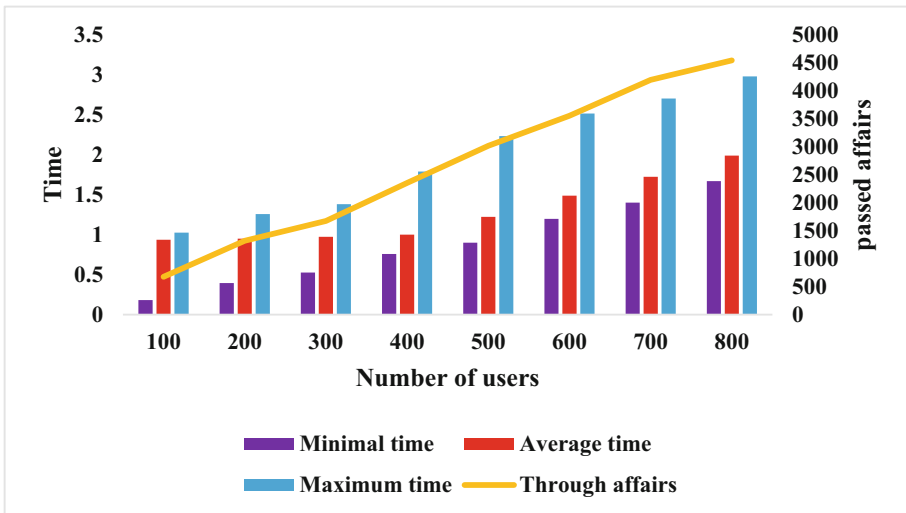
$$C = nL/T \tag{1}$$

$$C' = C + 3\sqrt{C} \tag{2}$$

In the formula, C is the average number of concurrent users, n is the number of login sessions, L is the average length of login sessions, and T is the length of the investigation time.

Table 1. Transaction response time

Transaction name	Concurrent users	Minimal time	Average time	Maximum time	Through affairs	Failed transaction
Q1	100	0.182	0.937	1.026	675	0
Q2	200	0.394	0.951	1.258	1321	0
Q3	300	0.526	0.973	1.382	1674	0
Q4	400	0.759	1.000	1.790	2356	0
Q5	500	0.900	1.224	2.234	3019	0
Q6	600	1.198	1.489	2.517	3555	0
Q7	700	1.402	1.726	2.704	4198	0
Q8	800	1.671	1.990	2.982	4545	0

**Fig. 2.** Transaction response time

The test results in Fig. 2 show that the response time of the system increases with the significant increase in the number of users. The current system allows up to 800 users to be used stably at the same time, meeting the system requirements for 600 people to visit and use, and the average response time is within 2 S. The maximum response time is also within 3 S required by the demand. After the above steps are executed, the test report shows that the system is running well and meets the requirements (Table 1).

5 Conclusions

By accelerating the construction of new infrastructure, we can expand domestic demand, increase employment, make up for the production gap caused by the epidemic, and offset

the economic recession. In addition, the appropriate promotion of new infrastructure construction will lay the foundation for long-term economic growth, significantly improve the efficiency of economic and social operations, and provide strong support for the long-term sustainable development of our country's economy.

References

1. Li, J., Chen, W., Chen, Y., et al.: A survey on investment demand assessment models for power grid infrastructure. *IEEE Access*, **99**, 1 (2021)
2. Shadare, A.E., Sadiku, M., Musa, S.M.: Electromagnetic compatibility issues in critical smart grid infrastructure. *IEEE Electromagn. Compat. Mag.* **6**(4), 63–70 (2018)
3. Chen, T., Cheng, Q., Li, X.: An anonymous key agreement protocol with robust authentication for smart grid infrastructure. *Sci. China Inf. Sci.* **65**(9), 1–3 (2021). <https://doi.org/10.1007/s11432-019-2736-5>
4. Yang, F., Zhang, S., Sun, C.: Energy infrastructure investment and regional inequality: evidence from China's power grid. *Sci. Total Environ.* **749**(2), 142384 (2020)
5. Fan, X., Li, Y., Luo, L.: Groundwater pollution detection and power grid enterprise human resource management based on Markov chain. *Arab. J. Geosci.* **14**(15), 1–16 (2021). <https://doi.org/10.1007/s12517-021-07751-7>
6. Hu, K.-Y., Li, W.-J., Wang, L.-D., Cao, S.-H., Zhu, F.-M., Shou, Z.-X.: Energy management for multi-microgrid system based on model predictive control. *Front. Inf. Technol. Electron. Eng.* **19**(11), 1340–1351 (2018). <https://doi.org/10.1631/FITEE.1601826>
7. Kumar, G., Nagarajan, C.: Design and implementation of intelligent energy control for tariff management using multi-agent system in smart/micro grid. *Concurr. Comput. Pract. Exp.* **31**(12), e4846.1–e4846.10 (2019)
8. Fiaschetti, L., Antunez, M., Trapani, E., et al. Monitoring and controlling energy distribution: implementation of a distribution management system based on Common Information Model. *Int. J. Electr. Power Energy Syst.* **94**, 67–76 (2018)
9. Nasr, M.A., Nasr-Azadani, E., Nafisi, H., et al.: Assessing the effectiveness of weighted information gap decision theory integrated with energy management systems for isolated microgrids. *IEEE Trans. Industr. Inf.* **16**(8), 5286–5299 (2020)
10. Srikantha, P., Kundur, D.: Hierarchical signal processing for tractable power flow management in electric grid networks. *IEEE Trans. Signal Inf. Process. Over Netw.* **5**(1), 86–99 (2019)
11. Han, D., Sun, W., Fan, X.: Dynamic energy management in smart grid: a fast randomized first-order optimization algorithm. *Int. J. Elect. Power Energy. Syst.* **94**, 179–187 (2018)
12. Kaur, D., Aujla, G.S., Kumar, N., et al.: Tensor-based big data management scheme for dimensionality reduction problem in smart grid systems: SDN perspective. *IEEE Trans. Knowl. Data Eng.* **30**(10), 1985–1998 (2018)



Risks and Countermeasures of Cyber Security Enterprises

Jianwei Ma¹, Zhuojun Yao¹(✉), and Sinem Alturjman²

¹ Department of Economics and Management, Lanzhou University of Technology, Lanzhou, Gansu, China

² Near East University, Nicosia, Turkey

Abstract. While the Internet penetration rate in China continues to increase, cyber security problems occur more frequently. The improvement of the level of risk management of cyber security enterprises is of great significance to the quality of the development of cyber security enterprises. Cyber security companies have the characteristics of fierce market competition, high R&D requirements, weak corporate management awareness, and large impacts from macro policies. Through public data analysis and the use of financial model of listed companies, it can be found that they face R&D risks, human resources risks, market risks, financial risks, and macro risks. Cyber security companies should strengthen product technology while paying attention to market demand, focusing on technical personnel and corporate management talents, flexibly choosing their own development strategies according to market conditions, and focusing on financial risks to reduce the above risks and achieve sustainable high-quality development. The government should increase support for cyber security companies and promote the importance of cyber security to promote the development of the cyber security industry.

Keywords: Cyber security enterprises · Risk management · Risk response

1 Introduction

Definitions of cyber security and cyber security risk differ between cyber experts and ontology developers [1]. Various factors, such as Socio-demographic characteristics, cyber security perceptions, cyber security breach experiences, IT usage, and knowledge influence security behavior [2]. With the development of cloud computing, big data, and mobile internet technology, cyber security are constantly being introduced, and mobile phone antivirus software, cloud security software, and virtual machine security software have emerged. Various cybersecurity companies are also transforming their product lines, hoping to find new opportunities. China is in a critical period of digital transformation. Few studies focus on the risks faced by cybersecurity enterprises, which is a key industry of digital transformation. From the perspective of enterprise risk management, we use data analysis and financial model to identify the risk of network security enterprises, and put forward countermeasures to promote the development of network security enterprises.

2 Characteristics of Cyber Security Enterprises

In September 2021, the Chinese cybersecurity media “Safety Bull” released the ninth edition of “China’s Top 100 Cybersecurity Companies”, QI-ANXIN, TOPSEC, HUAWEI, SANGFOR, New H3C, NSFOCUS, Alibaba Cloud, Venustech, and AsiaInfo are among the first echelons of the domestic cybersecurity industry. Analysis of them makes it easy to find the following characteristics of cybersecurity enterprises.

2.1 Great Sales Market Potential

The Internet penetration rate in China continues to increase. According to survey data in June 2021, the Internet penetration rate has reached 71.6%., and domestic Internet users have gradually realized the importance of cybersecurity, and the market demand for cyber security products will increase. Cyber attack is a new danger for critical infrastructures today [3]. Top management support, security policy, awareness/training key to building culture [4]. Under the guidance of the Chinese government’s cyber security strategy, downstream governments and corporate users have generally accelerated the construction of network informatization, and their investment in cyber security applications has increased (Figs. 1 and 2).

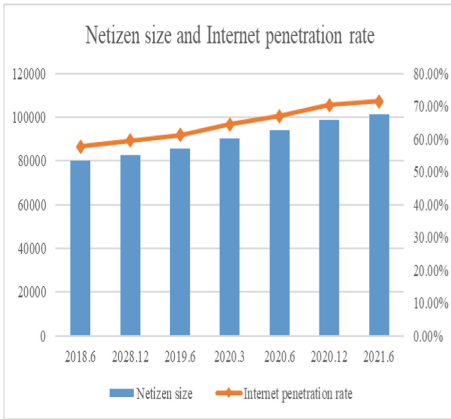


Fig. 1. Netizen size and Internet penetration rate

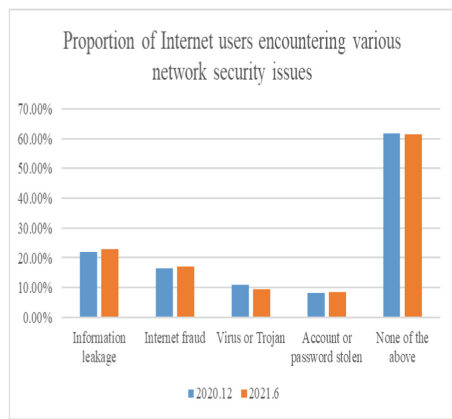


Fig. 2. Proportion of Internet users encountering various network security issues

2.2 High R&D Requirement

The cyber security industry is a knowledge-intensive industry. Compared with assets, technology is the most important resource of enterprises. The upstream companies mainly include manufacturers of mobile industrial computers, servers, integrated circuits and other related hardware equipment manufacturers, as well as software manufacturers such as operating systems and databases. Therefore, the market is fiercely competitive and the product technology is updated at a fast speed. Only by retaining R&D talents and continuously R&D investment can companies have a place in the fierce market competition.

2.3 Fierce Market Competition

The market share of cybersecurity products is gradually concentrated in enterprises with technical and brand advantages. Traditional cybersecurity products such as fire-wall, SOC, IDS/IPS, UTM have entered a mature period, and the market is relatively concentrated.

2.4 Weak Corporate Management Awareness

Cybersecurity enterprises are usually founded by technical personnel. The lack of knowledge of risk management by their leaders makes the entire enterprise relatively weak in awareness of risks. Rapid update technology and fierce competition make the risks unpredictable.

2.5 Affected by Macro Policies

Cybersecurity involves the storage and transmission of important data such as state secrets and commercial secrets. So they face the possibility of macro policy changes, such as Changes in preferential policies.

3 Risks Faced by Cyber Security Enterprises

3.1 R&D Risks

Table 1. R&D investment of network security enterprises

Co., Ltd.	Item	2016	2017	2018	2019	2020
TOPSEC	R&D	63.41	390	445.64	653.28	764.38
	Propotion	2.21	7.59	7.07	9.21	13.4
Venustech	R&D	429.76	469.67	547.33	604.58	657.99
	Propotion	22.3	20.61	21.7	19.57	18.04
NSFOCUS	R&D	231.41	289.88	317.29	337.54	365.43
	Propotion	21.22	23.1	23.59	20.2	18.18
SANGFOR	R&D	–	–	779.03	1140.89	1509.24
	Propotion	–	–	24.16	24.86	27.65
QI-ANXIN	R&D	–	–	–	–	1228.09
	Propotion	–	–	–	–	29.51

*R&D represents R&D investment and the unit is million yuan; Propotion represents the ratio of R&D investment to operating income and the unit is 1%.

Network security companies have the characteristics of limited public understanding, fierce competition, and rapid product updates. Cybersecurity industry often attach great

importance to the R&D. By looking up the data of listed cybersecurity companies, it can be seen that the R&D investment has increased year by year, and the patents applied by companies are mainly invention patents with high technical requirements, rather than using new or design patents. It can be seen from Table 1 that QI-ANXIN, which is China's strongest comprehensive cybersecurity company, has the highest R&D investment.

However, R&D is an activity with a great risk of failure. At the same time, market demand is difficult to accurately locate. This will lead to R&D risks.

3.2 Human Resources Risks

How long can a technology be developed and applied is uncertain, and alternative technologies will appear at any time. Enterprises must have a certain amount of technical manpower to support their continuous R&D activities. Therefore, the competition for technical talents (employment competition and lower turnover rate) is also the key for companies to compete.

One of the consequences of the high proportion of technical personnel is that managers often only pay attention to technical personnel, resulting in the lack of risk management personnel, and ignore the impact of risks on enterprise development (Table 2).

Table 2. R&D personnel of network security enterprise

Company name	Item	2015	2016	2017	2018	2019	2020
TOPSEC	Number	139	156	1418	1896	2265	2844
	Proportion	13.69	12.84	37.96	40.7	42.89	53.09
Venustech	Number	1206	1341	1412	1645	1936	2072
	Proportion	40.46	38.99	37.34	42.58	42	38.51
NSFOCUS	Number	982	1160	1315	758	811	928
	Proportion	55.64	59.92	59.99	27.85	27.58	27.28
SANGFOR	Number	–	–	–	1664	2247	3018
	Proportion	–	–	–	36.56	36.8	39.96
QI-ANXIN	Number	–	–	–	–	–	2899
	Proportion	–	–	–	–	–	37.19

3.3 Market Risks

The products provided by competitors are roughly the same. Some companies have expanded their production scale for the benefit of economies of scale and the products have been surplus. Companies have begun to resort to price-cutting auctions, forming vicious competition.

Meanwhile, domestic users are weak in genuine awareness, and the market is flooded with a large number of free products. Enterprises are facing great market uncertainty. Companies in the start-up stage may even face the risks of competitors completing the conversion of results first.

3.4 Financial Risks

Malicious competition has led to fierce competition, which has led to the failure to guarantee the profitability. Financial risks include, but are not limited to, insufficient cash flow, poor asset profitability debt servicing pressure. R&D activities have high input, high output risks, and face great financing constraints, which are likely to cause financial risks.

The Z-score model is used to evaluate the financial risks of network security companies [5], and the discriminant function is shown in formula 1.

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.999X_5 \quad (1)$$

Where $X_1 = \text{Working captical} / \text{Total assets}$

$X_2 = \text{Retained Earnings} / \text{Total assets}$

$X_3 = \text{EBIT} / \text{Total assets}$

$X_4 = \text{Market value equity} / \text{Total assets}$

$X_5 = \text{Sales} / \text{Total assets}$

$Z = \text{Overall Index}$

It can be seen that the Z-score of each company is greater than 2.67, there is no obvious sign of financial crisis. However, the gap between the companies' scores is large, indicating that the level of financial risk management between companies is quite different. In addition, the financial risk fluctuates greatly, indicating that the financial situation of is not stable, and financial risk management should be further strengthened (Fig. 3 and Table 3).

3.5 Macro Risks

Government support policies play a key role in cybersecurity companies, especially those in the early stages of development. The impact of policies on enterprises includes tax incentives, loan interest discounts, government special investment allocations, industry policies, government support plans, etc.

Table 3. Z-score

	TOPSEC	Venustech	NSFOCUS	SANGFOR	QI-ANXIN
2008	8.930384	–	–	–	–
2009	7.303651	–	–	–	–
2010	10.04993	7.780736	–	–	–
2011	12.83041	11.47686	–	–	–
2012	5.954245	6.543048	–	–	–
2013	3.840202	10.5809	–	–	–
2014	4.680154	11.10129	10.52736	–	–
2015	6.478246	19.71679	13.83304	–	–
2016	2.74748	14.51115	7.887836	–	–
2017	6.973639	14.85211	7.267152	–	–
2018	6.691057	12.13428	9.409932	4.921579	–
2019	10.34228	11.12523	21.26744	12.55953	–
2020	15.96727	11.43003	15.41434	19.40086	5.058342

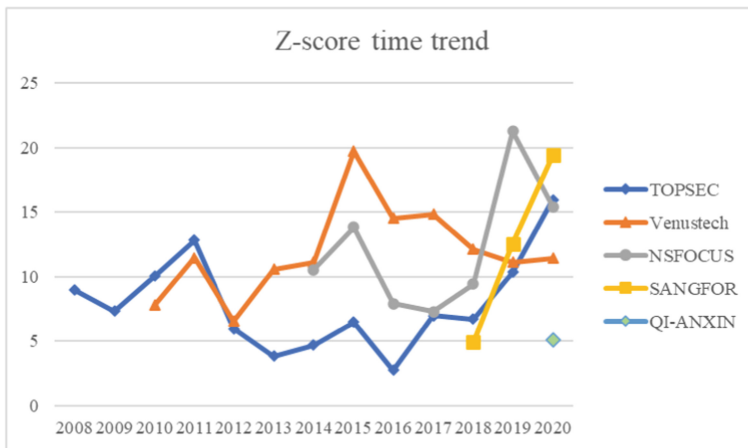


Fig. 3. Z-socre time trend

4 Cybersecurity Enterprises Risk Response Strategy

4.1 Strengthen Product Technology

Compared with foreign companies, domestic manufacturers must strengthen R&D in order to gain a place in the domestic market and even enter the international market. For enterprises that are lagging behind in technological innovation and development, they

can jointly develop new technologies with enterprises with mature and advanced technologies. We must also pay attention to customer needs, because the market addressed by the companies largely determines the process of innovation [6].

4.2 Carry Out Comprehensive Human Resource Management

The industry competition is actual the competition between industry technology and efficiency, while behind this competition is the competition of key talent [7]. Besides focusing on technical personnel, enterprises should pay attention to comprehensive management talents, including but not limited to financial risk management personnel, corporate development strategy planning talents, and marketing talents. For example, setting up a financing risk management department to analyze financing risks and apply corresponding risk aversion measures.

4.3 Have a Flexible Market Competition Strategy

First of all, for SMEs, the high-end market is mostly occupied by foreign companies or large domestic companies. But China is still in the period of industrialization. The demand from SMEs and low-income users will gradually increase. Cyber security SMEs should focus on the low-end market. Secondly, companies should choose a business model that suits them. Traditional cyber security companies only rely on malicious bidding on projects and use black-box operations to win orders. As a result, the profit margins have fallen and operating costs have increased. At present, some companies are transforming from cybersecurity product to cybersecurity services. Technical services are the development trend of network security companies [8], and companies obtain new profit growth points by providing customers with network security services.

4.4 Pay Attention to Financial Risk Management

R&D failure may lead to the rupture of the capital chain. Companies should assess the difficulty of R&D in advance, and choose an outsourcing strategy if it exceeds their own risk-bearing level. At the same time, companies should optimize their asset structure, and reasonably control the scale of investment and financing, so that cash flow can be coordinated in terms of time and amount. Secondly, companies should systematically evaluate financing risks, analyze the weak links in financing management, and make a risk contingency plan in advance by combing their own risk tolerance and preference and control risk factors with high probability of occurrence. This is also Actions in line with Enterprise Risk Management Framework [9].

4.5 Support Measures from the Government

The state should formulate corresponding policies to expand the financing channels of cybersecurity enterprises, because they are more difficult to operate than other enterprises. Due to large R&D investment and more capital demand, they are always faced with capital problems. In addition to using bank loans, the state should encourage investment in the field of cybersecurity and give preferential tax treatment. At the same time,

the government should publicize the importance of cyber security because firms' investments in information security are largely driven by external environmental and industry related factors [10].

5 Conclusion

With the proposal of the national security strategy, the government and enterprises have gradually realized the importance of cyber security. However, with the application of the Internet, cybersecurity problems such as information leakage, virus Trojan attack, network fraud and account theft occur frequently. It is urgent to promote the development of domestic cybersecurity enterprises. However, cybersecurity enterprises have their unique industry characteristics: 1. With the increasing market demand, there are many domestic and foreign enterprises in the industry, resulting in fierce market competition. 2. Putting R&D first is the foundation for enterprises, but it also leads to the dominant position of technicians and weak risk management. 3. The industry involves national security issues and is greatly influenced by macro policies. Through the analysis of the representative listed companies of China's cyber security enterprises, we find that cybersecurity enterprises face: 1. The R&D risk from high R&D technical difficulty, R&D failure and R&D products not meeting the market demand; 2. The risk of competition for technical talents and lack of professional management personnel; 3. Market risk from strong product substitution, fast renewal, malicious competition and market promotion failure 4. Financial risk from unstable financial situation and large gap in industry financial risk 5. Macro risk caused by changes in government policies. And. The following measures can help reduce the risks of network security enterprises: 1. Strengthen product R&D, reasonably evaluate the difficulty and cooperate with technologically advanced enterprises when necessary. Meanwhile, pay attention to market research. 2. Pay attention to enterprise management talents while attracting and reducing the turnover rate of R&D personnel. 3. Flexibly select market segments and business models in the fierce market competition and explore new profit growth points. 4. Make a systematic evaluation of financial risks and control weak links. Formulate emergency plan 5. The government should increase support for the cybersecurity industry and publicize the importance of it in order to improve China's cybersecurity technology level.

References

1. Cains, M.G., Flora, L., Taber, D., King, Z., Henshel, D.S.: Defining cyber security and cyber security risk within a multidisciplinary context using expert elicitation. *Risk Anal.* (2021). <https://doi.org/10.1111/risa.13687>
2. Kovacevic, A., Putnik, N., Toskovic, O.: Factors related to cyber security behavior. *IEEE Access* **8**, 125140–125148 (2020)
3. Gunes, B., Kayisoglu, G., Bolat, P.: Cyber security risk assessment for seaports: a case study of a container port. *Comput. Secur.* **103**, 102196 (2021)
4. Uchendu, B., Nurse, J.R.C., Bada, M., et al.: Developing a cyber security culture: current practices and future needs. *Comput. Secur.* **109**, 102387 (2021)
5. Edward, A.: Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Finance* **23**(4), 589–607 (1968)

6. Docter, J., Van Der Horst, R., Stokman, C.: Innovation processes in small and medium-size companies. *Entrep. Reg. Dev.* **1**(1), 33–52 (1989)
7. Guan, X., Li, M.: Information industry security: an industry convergence point of view. *Int. J. Inf. Technol. Manage.* **16**(3), 269–286 (2017)
8. Schneier, B.: The death of the security industry. *IEEE Secur. Priv.* **5**(06), 88 (2007)
9. Coso, I.I.: Enterprise Risk Management-integrated Framework. Committee of Sponsoring Organizations of the Treadway Commission (2004)
10. Weishäupl, E., Yasasin, E., Schryen, G.: Information security investments: an exploratory multiple case study on decision-making, evaluation and learning. *Comput. Secur.* **77**, 807–823 (2018)



Environmental Economic Decision Support Information System Based on GIS

Yuqi Wang^(✉)

Shanxi University of Finance and Economics, Taiyuan 030000, Shanxi, China

Abstract. Environmental and development issues are the focus of attention in the world today. The analysis of the impact of projects on the environment is included in the evaluation of investment projects, and sustainable development, not just economic benefits, is used as the evaluation criteria for investment projects, which is important for improving and promoting human society Comprehensive progress is of great significance. The purpose of this paper is to study the environmental economic decision support information system based on GIS. Firstly, it introduces the feasibility analysis, demand analysis project and the implementation of key GIS technology based on the theory of environmental economics. Secondly, it introduces the data collection and design of GIS environmental economic decision support information system, including data collection, data processing, and data organization and management mode. It is the realization of the auxiliary decision support system for city M urban planning based on GIS, including the introduction of the system and the operation of the system interface. Analyzing the Gaussian atmospheric diffusion model module and the statistical calculation module in the decision support system, the green volume of the E and F areas both reached 55%, which provides better data support for the planning and design of the city.

Keywords: GIS technology · Environmental economy · Decision support · Information system

1 Introduction

In recent years, with the development of computer technology, its application fields have been expanding, especially in the development of graphics and imaging technology, the application of data mining technology, and the expansion of economic management [1, 2]. People began to pay attention to the research and application of social media information sharing sites, usually various non-economic databases, that is, processing data that is not directly related to geographic data into GIS [3, 4]. GIS management tools have been further developed and improved in terms of scale and compatibility [5]. Judging from the current internal facilities, the distance between the application of local information systems in the business and economic fields and the facilities in other fields such as public planning and management, energy communications, land management, and environmental protection. With the establishment and improvement of the market economy, enterprises The improvement of analysis and strategic decision-making level and efficiency requires technical support that keeps pace with the times [6].

Sustainability research has evolved from a single focus on one aspect to multiple aspects of the triple bottom line at the same time. Arya P balances the environmental and economic sustainability of logistics. A model based on a system dynamics approach explains the simultaneous interaction of these factors. Simulation of this model helps managers of logistics functions to determine the scale of technology investment in order to achieve environmental efficiency without negatively affecting economic performance [7]. Martinkus N introduced a multi-standard decision support tool (DST) that includes economic, environmental and social indicators to assess the reuse potential of existing facilities as wood biorefineries. The economic site selection criteria are represented by the components of biorefinery operating costs that vary geographically and spatially. Environmental standards are the global warming potential of the supply chain, measured by the greenhouse gases emitted by raw material procurement, pretreatment and transportation equipment. In addition, user-defined overall indicator weights are used to adjust the importance of the three indicators, thereby changing the overall facility score. DST was applied to case studies in western Oregon and western Washington to refine the list of candidate pulp mills into a few for further investigation. It turns out that when the overall metric weight is adjusted, the factory scores are different. Therefore, different stakeholder preferences may produce different facility priority lists [8]. GIS will become an important application technology in this field in the future, and then form a decision support system for professional applications.

The research content of this paper mainly includes the following aspects: Based on the theoretical analysis of environmental economy and the discussion of key technologies, the establishment of an environmental economic decision-making system that supports geographic information. Research the application of GIS spatial analysis function in environmental economic decision support geographic information system. The spatial analysis function of the research application solves the distribution analysis of atmospheric environmental sensitivity, the application of public facility distribution overlay analysis, the multi-factor comprehensive evaluation and application analysis of characteristic entities, and the optimization of commercial network layout and location selection.

2 Research on Environmental Economic Decision Support Information System Based on GIS

2.1 Environmental Economics

The environment is a resource, not the wealth that nature has long known to humans [9, 10]. It provides raw materials for human production activities, including renewable and non-renewable resources such as land, water, forests, and minerals, and is the material basis for economic development. That is, the habitat that humans depend on can diffuse, store and assimilate the waste generated by human activities, and provide services for the human landscape, which is the basic condition for the survival of the tourism industry [11, 12].

Environmental resources have the following characteristics:

(1) Scarce environmental resources

Although the total amount of most environmental resources is relatively abundant, the stock is always limited. In terms of social needs, environmental resources have become scarce.

(2) Environmental resources can be reused

For specific environmental resources, its use is multifaceted. For example, as a resource lake, its uses include freshwater aquaculture, industrial water, agricultural water, fishery areas, and waterways. Some of these uses are mutually exclusive and cannot be used for multiple purposes at the same time. This requires full attention to the comprehensive utilization of resources in the development and utilization of environmental resources, and strive to maximize benefits.

2.2 GIS Technology

- (1) The basic functions of GIS map and imaging determine seamless data integration and seamless map optimization. Layer display control, switch level and eagle eye window to determine the approximate position of the window on all maps when navigating the map. Click “Search Map Properties” to get project information, measurement tools, position detection, length measurement, object area, and display size. In addition, there are graphic editing, image adding/deleting, template and conversion function editing, etc.
- (2) GIS search/analysis function/search map attributes and enter topics such as place names to identify the location of the attributes on the map. Analyze the shortest route and determine the nearest facility and service area and other services. The function of storage analysis is to find feature information in the scale from the scene, and it also includes common GIS spatial analysis services.
- (3) The monitoring service uses cameras and sensors to monitor various risk sources in real time and record monitoring information. When an environmental threat occurs, you can be notified in time, notify the corresponding environmental emergency department, and quickly and accurately locate the alarm on the map.
- (4) The environmental model library system needs to provide access to environmental pollution models and integrate GIS functions with the model library. For example, after an environmental risk occurs, you can use the pollution and damage calculated by the environmental pollution model to draw the results or use the GIS function to create a corresponding map.
- (5) For the stability and security of data management, the spatial database must be performed by the spatial data engine.
- (6) The created thematic map, text and preview function can be edited before printing.

3 Investigation and Research on Environmental Economic Decision Support Information System Based on GIS

3.1 Introduction of GIS Environmental and Economic Decision Support Information System

The GIS-based urban environmental economic decision support system of M City is based on the MapGIS10 platform using C# language for secondary development, and provides environmental and economic decision support for the urban construction of M City. The appearance design of this system pays attention to the friendly interface, the reasonable layout, the clear function and the convenience of operation, so as to improve the efficiency of the system as much as possible and reduce the work intensity. The function setting is compact and practical, while ensuring accuracy and precision, which is convenient for system users to query location information, house attribute query and corresponding analysis. It can also perform more advanced statistics based on system data to assist system users in making decisions.

The hardware environment of the system is mainly server and client. The specific requirements are as follows:

Server: PC server, hard disk above 100 G, memory above 512;

Client: The main configuration requires CPU 2.4 Hz or more, memory 215 M or more, independent graphics card.

The software environment for system operation mainly includes three parts: operating system, database management system and software support environment. The specific requirements are as follows:

Operating system: mainly Windows XP;

Database: Oracle9i and above products;

Supporting environment: mainly Directs X.

3.2 Data Collection

The original data used in the surface model data is satellite image data, DEM data, and data collected on the spot using a total station. The texture photos of all models are collected on the spot, and finally fine models and batch models are constructed from these data. The satellite image uses the image map of the Gaofen 2 satellite in the urban area of Yiyang City as the original image of the image, and then through a series of image processing such as radiometric calibration, atmospheric correction, orthorectification, image stitching, image fusion, image cropping, etc. Standard image map.

3.3 Data Preprocessing

The acquired images have been subjected to radiation correction and rough geometric correction at the remote sensing satellite ground station, but there are still many unsystematic errors, such as image geometric distortion caused by factors such as terrain undulations and inaccurate ephemeris data, so geometrical corrections are required. Fine correction.

The data is spatially transformed according to the control points. The relationship between the number of polynomial powers and the minimum number of control points is shown in formula 1:

$$(t + 1)(t + 2) / 2 \quad (1)$$

Among them, t is the power number. When the collection of all control points is completed, the bilinear interpolation method is used to resample the image of the collected control points, and finally a corrected remote sensing image is obtained.

After the pollutants are discharged into the atmosphere, they will undergo a series of physical and chemical processes to gradually reduce their concentration. The Gaussian diffusion model of elevated continuous point sources is shown in formula 2:

$$C(x, y, z, H) = \frac{Q}{2\pi u_x \sigma_y \sigma_z} \left[e^{-\frac{1}{2} \left(\frac{y^2}{\sigma_y^2} + \frac{(z-H)^2}{\sigma_z^2} \right)} + e^{-\frac{1}{2} \left(\frac{y^2}{\sigma_y^2} + \frac{(z+H)^2}{\sigma_z^2} \right)} \right] \quad (2)$$

In the model, $C(x, y, z, H)$ is the pollutant concentration at coordinates (x, y, z) . H is the effective source height.

4 Investigation and Research Analysis of Environmental Economic Decision Support Information System Based on GIS

4.1 Gaussian Atmospheric Diffusion Model Module

The Gaussian atmospheric diffusion model module in the decision support system adopts the VS C# development language, uses the digital map of M city as the base map of the system, and uses the specific model framework as shown in Fig. 1. Enter each parameter value of the pollutant in the environmental assessment support system, determine and enter the actual weather condition parameters, realize the Gaussian atmospheric diffusion model through visual C# programming, and continuously calculate the model within a certain range of the downwind direction, and import the large amount of data generated into the database middle. The data is automatically imported through the interface between the database and the Golden software software, and the contour map of the pollutant concentration is drawn in Surfer. The system overlays the final contour map of pollutant concentration and the digital city map of M city, thus providing a visual basis for decision support.

The evaluation of the final atmospheric pollutant concentration results should consider not only the pollution level of the actual concentration value, but also the specific requirements of the environmental functional zoning of City M. By calling the thematic maps of different functional zoning in the digital map, the environmental assessment results can be more intuitive, and convenience.

Comprehensive analysis of the current status and development trend of the social functions of each environmental unit, the distribution of atmospheric environmental sensitivity, and the comprehensive zoning results of the urban ecological environment,

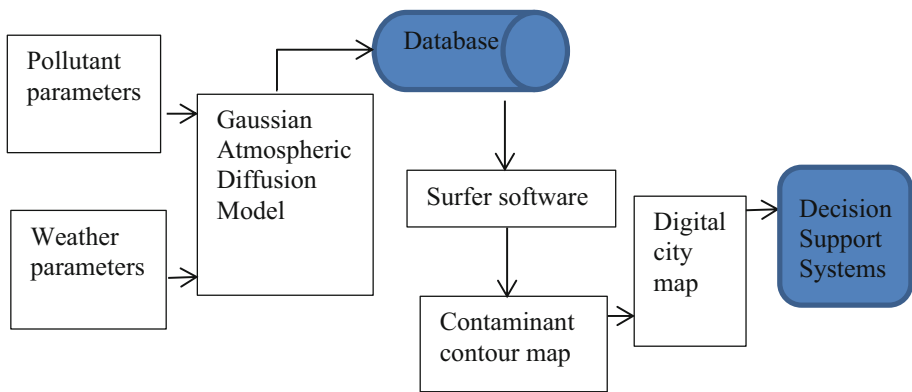


Fig. 1. A schematic diagram of the framework of the Gaussian Atmospheric Diffusion Model Module

determine the dominant function of each environmental unit, according to the environmental air quality requirements of the protected object, and Ambient air quality functional areas are divided into three categories: the scope of each area, leading functions and implementation standards.

The environmental quality functional zoning of City M is formulated in accordance with the actual environmental protection requirements and standards of City M. Its function takes into account the actual division of business districts, residential areas, cultural and educational districts in City M, and the environmental protection standards formulated and implemented are more strict.

4.2 Statistics Calculation Module

The statistical calculation function module realizes the two functions of selecting objects and attribute statistics. These two functions are used to perform statistics and analysis on the primitives and layers in the map document, which is convenient for system users to make decisions based on the statistical analysis results, so as to achieve the purpose of system-assisted decision-making.

You can select the map interface for data statistics and tabulation, and you can also perform on-site graphic element measurement as shown in Table 1. Element statistics is mainly to grasp some basic data of the city, and provide better data support for the planning and design of the city. The current situation of the city can be grasped from the overall city as shown in Fig. 2. Which public facilities are still too few and where are unevenly distributed, so that these factors will be taken into account in the future design.

Pixel measurement is mainly based on some design factors of the building. How big is the area, what is the floor area ratio, and what is the greening rate. These information can be calculated through our pixel measurement.

Table 1. Element statistics

Area	Building facilities	Greening facilities	Other
A	60%	32%	8%
B	50%	45%	5%
C	66%	28%	6%
D	70%	28%	2%
E	45%	55%	5%
F	40%	55%	15%
G	56%	34%	10%
H	65%	20%	15%

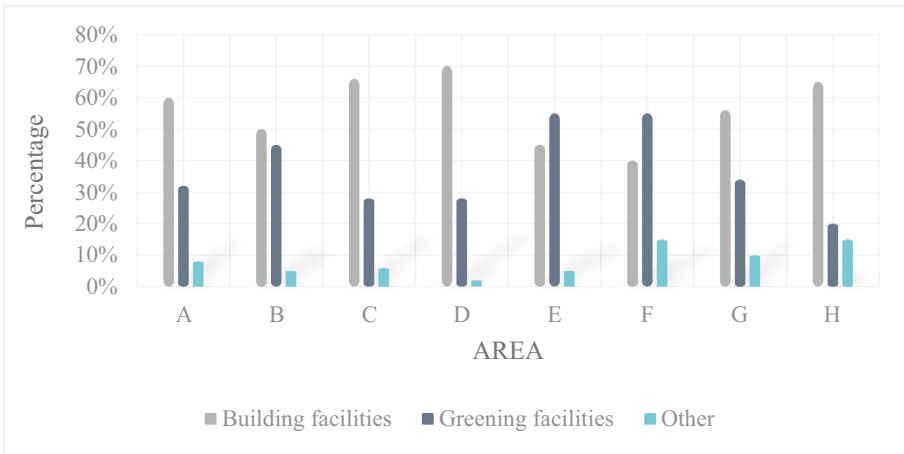


Fig. 2. Feature statistics

5 Conclusions

Information and computing have intensified the process of globalization, shaped new production methods and business rules, and of course also affected the strategy, planning, and implementation of regional economic development. The interactive development of environment and economy is an inevitable requirement of modern society. Ecological construction and economic interactive development are the combination of ecological and economic knowledge to promote social and economic development. Based on the research of geographic information system technology, remote sensing technology and other spatial information technology, combined with the development of M City, the application of geographic information and geographic information technology in the enterprise and regional economy was systematically analyzed, and an environmental economic decision-making support geographic information was constructed. System, improve the regional business environment, enhance the competitiveness of the

regional economy, shape the regional image, and create conditions for the sustainable development of the regional economy.

References

1. Yang, L., Wang, H., et al.: Analysis of endemic situation of schistosomiasis based on geographic information system (GIS) in Wuhan City in 2017. *Zhongguo xue xi chong bing fang zhi za zhi = Chin. J. Schistosomiasis Control*, **31**(4), 410–413 (2019)
2. Tanjung, M., Syahreza, S., Rusdi, M.: Comparison of interpolation methods based on geographic information system (GIS) in the spatial distribution of seawater intrusion. *J. Nat.* **20**(2), 24–30 (2020)
3. Li, J., et al.: A two-stage water allocation strategy for developing regional economic-environment sustainability. *J. Environ. Manage.* **244**, 189–198 (2019)
4. Chaloob, I.Z., Ramli, R., Nawawi, M.: Measuring economic and environmental efficiency for agricultural zones in Iraq using data envelopment analysis. *Int. J. Inf. Decis. Sci.* **10**(3), 235–248 (2018)
5. Umer, A., Hewage, K., Haider, H., Sadiq, R.: Sustainability evaluation framework for pavement technologies: an integrated life cycle economic and environmental trade-off analysis. *Transp. Res. Part D: Transp. Environ.* **53**, 88–101 (2017)
6. Waldick, R., Bizikova, L., White, D., et al.: An integrated decision-support process for adaptation planning: climate change as impetus for scenario planning in an agricultural region of Canada. *Reg. Environ. Change* **17**(1), 1–14 (2017)
7. Arya, P., Srivastava, M.K., Jaiswal, M.P.: Modelling environmental and economic sustainability of logistics. *Asia-Pac. J. Bus. Administration* **12**(1), 73–94 (2019)
8. Martinkus, N., et al.: A multi-criteria decision support tool for biorefinery siting: using economic, environmental, and social metrics for a refined siting analysis. *Biomass Bioenergy* **128**, 105330 (2019)
9. Lechner, A.M., McIntyre, N., Witt, K., et al.: Challenges of integrated modelling in mining regions to address social, environmental and economic impacts. *Environ. Modell. Software* **93**, 268–281 (2017)
10. Gibson, J.: Forest loss and economic inequality in the solomon islands: using small-area estimation to link environmental change to welfare outcomes. *Ecol. Econ.* **148**, 66–76 (2018)
11. Basic-Sontic, A., Czap, N.V., Fuerst, F.: The role of personality traits in green decision-making. *J. Econ. Psychol.* **62**, 313–328 (2017)
12. Singh, A., Basak, P.: Economic and environmental evaluation of municipal solid waste management system using industrial ecology approach: evidence from India. *J. Cleaner Prod.* **195**, 10–20 (2018)



Network Security Management and Protection in the Context of Emerging Technologies

Qian Li^(✉)

Department of Management, Wuhan University of Technology, Wuhan, Hubei, China

Abstract. Digital technologies have expanded the reach of financial services to end-users, improved efficiency and driven global financial innovation to a new level. It has created new directions and opportunities for the transformation and development of the financial industry, but at the same time it has also brought new problems of network security management. In this paper, we study the cybersecurity issues of the financial industry in the context of emerging technologies. The results find that the high-benefit, regulated and high-risk characteristics of the financial sector have led to the prominence of financial data security issues, a significant increase in cyber attacks and more insidious cyber fraud in the financial sector. To this end, it is proposed to implement the network security ecological assessment system as soon as possible, to conduct regular network security attack and defense drills, to strengthen the application of technology in network security, to focus on key areas to carry out publicity activities and other protective measures to build a solid financial network security “firewall”.

Keywords: Emerging technologies · Network security · Protection · Financial industry

1 Introduction

Artificial intelligence technology has made breakthroughs. The development of a new generation, represented by artificial intelligence, has brought new growth opportunities to the financial industry. Naturally, it causes to undergo radical changes. The financial industry has evolved from manual operations to electronic operations in the second half of the 20th century, then information technology at present, and will develop in the direction of intelligence in the future (see Fig. 1). Digital technology has changed the scope for financial services to reach end users [1]. However, opportunities often go hand in hand with risks. This trend has dictated the new situation in which network information security has come into being.

2021 is the 100th anniversary of the founding of the CPC and the first year of the 14th Five-Year Plan. Under the complex and changeable security environment, the State has not only continuously raised the attention and importance of society to cybersecurity from the legislative level, but has also introduced a number of cybersecurity-related policies one after another. On March 11, 2021, the “Outline of the 14th Five-Year Plan

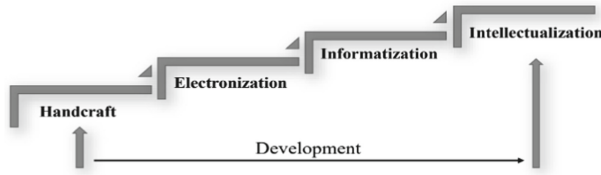


Fig. 1. Diagram of the evolution of the financial sector

and Vision 2035” clearly proposed “strengthening network security protection, enhancing network security infrastructure construction and improving network security threat detection”. The People’s Bank of China also issued Measures for the Administration of Credit Business on 11 January 2021, stating that it will establish dedicated departments to manage information security and inspect the implementation of security management systems regularly. All of those reflect the importance China attaches to cybersecurity and will accelerate the process of building cybersecurity infrastructure.

However, the security posture of the financial industry is still getting more and more serious. Cyber security and data leakage incidents are continuing to make headlines. The issue of cybersecurity in the financial industry is in urgent need of protection. Therefore, this paper combines the characteristics of the financial industry, explores the problems faced by network security and proposes targeted protection measures aiming to build network information security. The research framework is shown in Fig. 2.

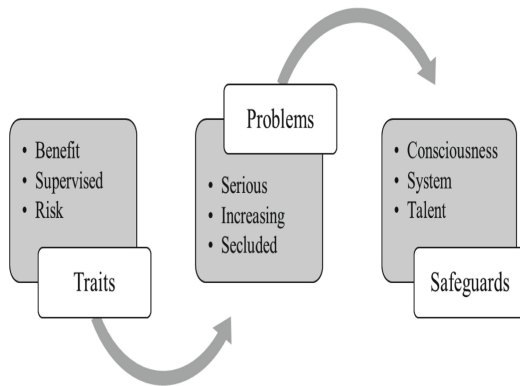


Fig. 2. Research framework

2 Characteristics of Financial Industry in the Context of Emerging Technologies

2.1 High Benefits

Based on intelligence and online services, the scope of services in the financial industry has been expanded. Restricted by cost and risk, traditional financial services are only

for large financial customers. Big data technology has provided more social individuals with access to financial services [2, 3]. As in the case of Balance Pay, it allows more small users to enjoy higher interest rates on their deposits on a pay-as-you-go basis.

2.2 Supervised

Throughout the world, all countries with a market economy system have the commonality of government regulation of the financial system [4]. In recent years, with various new types of financial services flocking to the market, the trend of digital economy is unstoppable. And new risks have emerged especially in the rise of internet finance such as P2P and microfinance. Therefore, the regulatory systems need to keep pace with the digital economy. Whether it is major banks offline or online payment platforms such as Alipay and WeChat, they all need to be connected to the national Netlink and regulated by the central bank.

2.3 High Risk

The financial industry's shift from a single business model to a globally networked and diversified model has brought new challenges to financial information security. When market transparency accelerates, the financial structure also changes from a single risk to a dual risk. As a result, China's fintech has also shifted from developing markets to preventing and maintaining financial security. The financial sector holds a large number of customer information. Thus, the risk of leakage is incalculable.

3 Cybersecurity Issues in the Financial Industry in the Context of Emerging Technologies

3.1 Data Security Problems in the Financial Industry Have Come to the Fore

The rise of emerging Internet technologies has not only brought transformation opportunities to the financial industry, but also brought opportunities for lawless elements to take advantage of. About 10% of data breaches are in the financial sector. The average financial industry has 352,771 exposed sensitive files. And nearly two-thirds of financial companies have more than 1,000 sensitive files open to each employee. With an average of 233 days for the financial industry to detect and address a data leakage, coupled with its high cost, which averages about US\$5.85 million, the financial industry's data security problems are highlighted.

3.2 Significant Increase in Cyber Attacks in the Financial Sector

The expansion of online transactions and contactless payments by financial institutions has been accompanied by an increasing activity of cyber attackers, posing a serious threat to personal privacy and financial security. In recent years, a number of national hacking groups have grown rapidly, taking advantage of cyber attacks to steal intelligence and compromise national security and this figure continues to climb. 2021 saw significant losses. Significant cyber security incidents in the global financial sector are shown in Table 1.

Table 1. Major cyber security incidents in the global financial sector in 2021

Time	Events	Causes
Late March 2021	US insurance giant CNA spends \$40m to decrypt data (the highest ransom)	Data stolen in ransomware attack
2021.8.19	Crypto assets worth \$94 million stolen from Japanese cryptocurrency exchange Liquid	Hot wallet under cyber attacker's control
2021.9.2	Russian banking blows up with access glitches	Massive DDoS attack
2021.9.13	First attack on the country's core financial system (Brazilian Treasury)	Ransomware attacks
2021.10.10	Ecuador's largest private bank suffers cyber attack, widespread business disruption	Ransomware attack with Cobalt Strike beacons installed on banking networks
2021.10.30	NBP forced to discontinue related services	NBP under cyber attack
2021.11.1	Jewellery giant Graff attacked by ransomware Conti	Conti gang claims to have stolen 69,000 confidential documents

3.3 Cyber Fraud in the Financial Sector Has Become More Insidious

The financial industry has become the most affected area in terms of online fraud. In the context of emerging technologies, people's lives are transformed to online consumption and new types of online fraud continue to be prevalent including impersonating financial institutions' customer service fraud, phishing software fraud, Jitterbug video fraud, joint code fraud, e-commerce refund fraud, campus loan cancellation fraud, fake shopping fraud, part-time job fraud, etc. These network frauds are more covert, more fraudulent and even take the strategy of "catching big fish with long lines", which is difficult for the general public to detect if they are not strong in fraud prevention.

4 Management and Prevention of Network Security in the Financial Industry in the Context of Emerging Technologies

4.1 Implement the Network Security Ecological Assessment System as Soon as Possible

Currently, under the guidance of the new development concept, actively promoting the digitization of industries and guiding the deep integration of the real economy can help promote high-quality economic development [5]. Cybersecurity issues in the financial sector, which are fostered by the emerging technology sector, can be tackled in a digital technology way to achieve multiplier effort. Therefore, with digital economy as the guide, penetrate financial cybersecurity supervision information, integrate on-site supervision

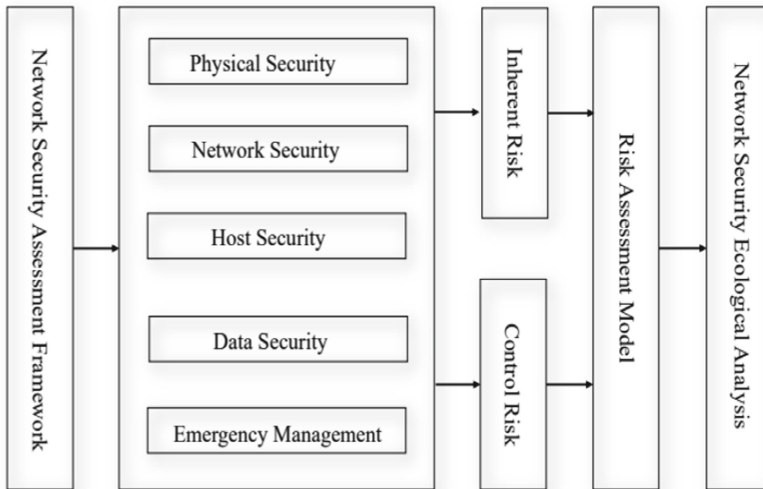


Fig. 3. Cybersecurity ecological assessment model for the financial sector

and off-site supervision information effectively, promote the digitization of various types of information vigorously, establish a sound mechanism for the digital transformation of supervision information, allow full play to the supporting role of digital in supervision and decision-making, establish a financial industry cybersecurity ecological assessment model (see Fig. 3) and conduct early cybersecurity warning.

4.2 Conducting Regular Network Security Defense Drills

Cyber defense drills are mainly used to simulate hacking attacks at home and abroad and to conduct penetration tests on an organization's network platforms and information systems, with the aim of identifying vulnerabilities that can lead to data leakage, asset damage and system tampering. Therefore, it can provide a chance to help the organization to identify security risks early in maintaining cyberspace security [6, 7]. The introduction of cyber defense drills into the financial sector is of great relevance and long-term value, especially for emergency management of financial infrastructure. Its own characteristics have the value of a "PDCA" cycle of improvement for overall emergency management from an IT perspective (see Fig. 4).

4.3 Strengthening the Application of Technology in Cybersecurity

Under the innovative development trend of big data, artificial intelligence, cloud computing and other emerging technologies, the threat of cyber security in the financial industry is more extensive, more sudden and more destructive. In the face of security risks, traditional regulatory measures are slightly backward and lack effectiveness and timeliness [8, 9]. And technological means should be fully applied to the whole process of pre-business, in-business and post-business management to expand the coverage of cyber risk protection and enhance the ability to screen which can prevent and resolve systemic and regional cyber security risks in Hull.

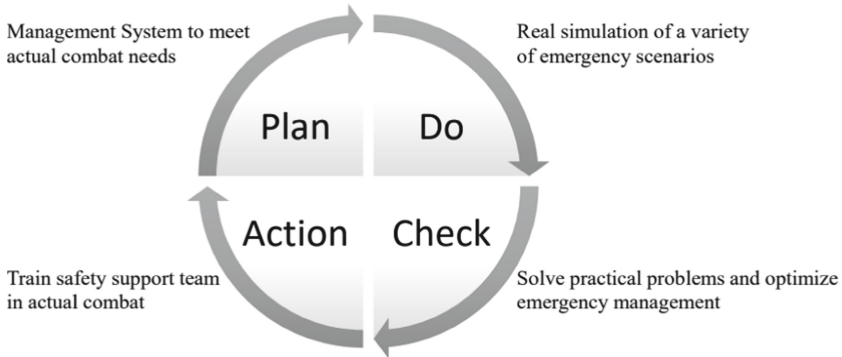


Fig. 4. “PDCA” cycle of the network security attack and defense exercise

4.4 Focus on Key Areas and Carry Out Promotional Activities

Publicity points were set up in front of the business branches to carry out focused publicity activities to densely populated areas such as markets, schools and surrounding communities. And organizing specialized staff to promote and popularize knowledge of cyber fraud prevention for customers handling business and passing citizens. Taking into account the relatively weak awareness of financial security among the elderly and new students, students and elderly groups were positioned as special groups for focused publicity. Conduct special training jointly with relevant financial institutions [10]. At the same time, the publicity campaign was close to the rural areas, farmers and the actual situation. Carry out in-depth front-line publicity in villages and enterprises to provide a real lesson in cyber security for the rural grassroots people. By analyzing common current fraud cases and explaining how to spot a scam, focus on maintaining personal financial network information security and protection from illegal fraudulent activities and other related financial knowledge to improve the public’s awareness and skills in preventing financial network fraud.

5 Conclusions

In order to effectively deal with technical and business risks in the financial industry, it is necessary to be vigilant about the spread of risks, pay more attention to improving regulatory measures in accordance with the law. Enhance the ability to identify cyber security incidents, strengthen the application of technology in cyber security and coordinate the work of key groups in cyber security according to the coordination mechanism of financial technology. On this basis, build a solid “firewall” for financial cyber security.

References

1. Yan, W.F., Shu, Q., Gao, P.: Security risk prevention and control deployment for 5G private industrial networks. *China Commun.* **18**(9), 167–174 (2021)

2. Nam, K., Seong, N.: A study on influencing factors for customer satisfaction and the continuing use of social network services in financial industry. *Enterp. Inf. Syst.* **15**(3), 395–419 (2020)
3. Chalmers, A.W., Young, K.L.: Network effects in the formation of the financial industry's regulatory preferences in the European Union. *Bus. Polit.* **22**(1), 52–84 (2016)
4. Minna, F., Chandrasekaran, B., Blaise, A., Rebecchi, F., Massacci, F.: Understanding the security implications of kubernetes networking. *IEEE Secur. Privacy* **19**(5), 46–56 (2021)
5. Zhang, M.Y., Wang, L.Y., Jajodia, S., Singhal, A.: Network attack surface: lifting the concept of attack surface to the network level for evaluating networks' resilience against zero-day attacks. *IEEE Trans. Dependable Secure Comput.* **18**(1), 310–324 (2021)
6. Wang, X.J., Gan, L.S., Liu, S.L.: Research on intelligence analysis technology of financial industry data based on genetic algorithm. *J. Supercomput.* **76**(5), 3391–3401 (2020)
7. Lin, Y.Z., Abur, A., Xu, H.C.: Identifying security vulnerabilities in electricity market operations induced by weakly detectable network parameter errors. *IEEE Trans. Industr. Inf.* **17**(1), 627–636 (2021)
8. Wang, L.G., Yu, J.Y., Qiao, B.J.: Intelligent evaluation of computer network security based on neural network. *J. Intell. Fuzzy Syst.* **35**(3), 2887–2891 (2018)
9. Lu, X., Niyato, D., Privault, N., Jiang, H., Wang, P.: Managing physical layer security in wireless cellular networks: a cyber insurance approach. *IEEE J. Sel. Areas Commun.* **36**(7), 1648–1661 (2018)
10. Chae, H.K., Jooyeoun, L.: Research on major technology trends in the field of financial security through Korea and foreign patent data analysis. *J. Digital Convergence* **18**(6), 53–63 (2020)



Network Security Design of E-commerce Sales Management System Based on Neural Network Algorithm

Shaocheng Zhang, Jiali Yang, Tingwei Wang, and Cong Wang^(✉)

School of English Language and Literature, Xi'an Fanyi University, Xi'an 710105,
Shaanxi, China
wangcong@xafy.edu.cn

Abstract. With the continuous development of the e-commerce marketing management system and the expansion of the network scale, the security of the e-commerce marketing management network is under more and more threats, and the attacks are becoming more and more serious. Frequent network security incidents have caused huge losses to the e-commerce marketing management system, and a variety of network security technologies need to be used to solve them. The purpose of this paper is to study the network security design of e-commerce sales management system based on neural network algorithm. This paper designs a complete network security intrusion detection system, from the design of each module to the design of the entire system. The system uses the optimized neural network as the intrusion detection algorithm, and tests the detection accuracy of the overall data detection module. Finally, the development of the entire system is completed, and the functional tests of the various modules of the system are carried out. Experimental research shows that the predicted value of the simulation experiment results in this paper is very close to the actual risk assessment results, and the error between the two is less than 2%.

Keywords: Neural network algorithm · E-commerce sales management system · Network security · Intrusion detection

1 Introduction

With the rapid development of e-commerce sales management systems, network threats such as malicious network attacks and information theft continue to threaten people's lives and work [1, 2]. With the increase in the number of Internet users, the amount of data in the e-commerce marketing network also increases. The Internet has the characteristics of openness, connectivity and sharing. Therefore, the complexity and unpredictability of the network are getting higher and higher, and it is becoming more and more obvious, and the problem of safe network is also unfolding.

In the research on the network security design of the e-commerce sales management system based on neural network algorithms, many scholars have studied it and achieved good results. For example, Soni N provided more information about network security

knowledge and defined an appropriate the model can detect and prevent attacks in a defined network environment [3]. Awad A uses virtual networks to study network security knowledge, design modular technology and related processes [4]. It can be seen that the research on computer system network security based on neural network algorithm is of great significance.

This article describes the related design of intrusion detection system. The program uses a virtual network as an intrusion detection algorithm and checks the accuracy of the general data search module. This article introduces the traditional network security model and its related improvements. The improved model may be more suitable for evaluating and predicting network security status.

2 Research on Network Security Design of E-Commerce Sales Management System Based on Neural Network Algorithm

2.1 Application of Bp Neural Network in Intrusion Detection System

(1) Intrusion detection process

1) Information collection

The first step in detecting intrusions is data collection. Data collection refers to the collection of information received from key areas of the network system, including the location and behavior of the system, network, data, and user services. Including network database, network log, directory and external file modification, system scan log, running external system, etc.

2) Information preprocessing and analysis

Data extraction refers to the preliminary collection of data and transforming it into a format and content that researchers can recognize. When a fault is detected, an alarm will be generated and sent to the console to determine whether an intrusion has occurred [5, 6].

3) Respond according to the security policy

The program responds accordingly based on the predetermined response generated by the alarm. The answers given here refer to active security measures, such as starting search traffic, warning the controller, disconnecting the network, or changing firewall settings.

(2) Divided according to the detection model used by the detection system

1) Use model anomaly detection

The basic premise for finding anomalies is that the behavior of the intruder is very different from the behavior of the normal object. By adjusting and recording

system functions, and then analyzing the monitoring data, a data set of the main physical characteristics of the system is established. Match the current state of the program with the normal behavior library [7, 8]. If there is a large deviation, then the service is intrusive. The advantage of intrusion detection method is to detect intrusion by detecting intrusion behavior, and there is no requirement for intrusion characteristics. However, it is difficult to solve problems such as setting behavior index values, setting normal behavior index values, and reducing the false alarm rate.

2) Use the abuse detection model

The basic premise of intrusion detection is to record and analyze various types of hacking behaviors and establish a database of intrusion behaviors. Through this database source and detection of intrusion behavior, once the match is successful, it means that the intrusion has occurred. The advantage of this detection method is that it has a low false alarm rate and can detect and warn of known intrusions early [9, 10]. However, the search function of this method is completely affected by the integrity of the corresponding data intervention, and all databases must be updated regularly. In addition, this method cannot detect unknown intrusions.

(3) Divided according to the data source of the test

Intrusion detection systems are divided into three categories based on search data sources: server intrusion detection systems, network intrusion detection systems and distributed intrusion detection systems.

1) Network-based intrusion detection system

Network-based intrusion detection systems usually use message passing or process mapping to set rules. By downloading the data packet database on the Internet and calculating and analyzing all these databases, potential databases are obtained. Match these databases with the standard framework, and judge whether there is an intrusion based on the compatibility result.

2) Host-based intrusion detection system

Server-based intruders detect hacker intrusions by monitoring and analyzing the control files of the server. It integrates the search engine into the system to ensure security, first extracts system data from the system to ensure security, and then analyzes the test files to determine whether an intrusion has occurred.

3) Distributed intrusion detection system

Recently, the discovery of distributed intrusion has attracted more attention. The Distributed Access System (DIDS) collects and analyzes the data, and then processes the data in a distributed manner [11, 12]. The intrusion detection system may respond well. Among them, agent-based intrusion detection is listed as the

main research content. The data collected through DIDS can come from various host programs, network components, or DIDS can also use network monitoring to collect data.

2.2 Intrusion Classification of Command Automation Network Based on Neural Network

(1) Data acquisition module

The first step to detect intrusion is data collection. By setting downloads at each key point of the command automation network, collect all information flows, user activity information and location data flowing through the command automation network.

Data retrieval requires real-time operation. If the data in the network cannot be retrieved in time, the retrieval result will be invalid. The data collected must be complete. If the collected data is incomplete and cannot cover as much basic meeting information as possible, the search results will be reduced.

(2) Data preprocessing module

Intrusive network access systems must write data over and over again, so it is important to ensure that the data collection is complete and the best training can be performed. However, the model system always contains some rough data. Therefore, there must be the following content before detecting data: data processing. The data flow that is automated through the automation network is huge, and redundant data will reduce the training speed and cannot meet the needs of real-time search. In this way, incomplete data can be deleted, the efficiency of the search engine can be reduced, and the efficiency can be improved. The format of the data received from the command automation network is incompatible. Therefore, it is necessary to create content, further filter data, extract basic information, and filter unnecessary information. Analyze the data obtained after preprocessing, and use feature matching to distinguish normal and abnormal behaviors.

(3) Intrusion analysis module

The intrusion analysis module uses analysis methods to analyze the information received from the command automation network. Commonly used methods include statistical analysis, shape mapping and full analysis.

Statistical analysis: Based on statistics, system objects such as files, processes, and users are described. Essentially, the configuration of these objects under normal operating conditions is calculated, including access frequency, time delay, service failure rate, etc. Compare the average value of the parameter system with the known network. If the value exceeds the allowable average value, it indicates that the intrusion has occurred.

2.3 Normalization

Let x_{\max} and x_{\min} represent the maximum and minimum values in the sample data set respectively, x is the i -th sample data in the sample data set, and y is the normalized

result of the sample data. The formula of the first normalization method is:

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

After normalizing the sample data by formula (1), the data is mapped in the interval [0,1].

The second normalization formula is:

$$y = 2 \times \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} + (-1) \quad (2)$$

After normalizing the sample data by formula (2), the data is mapped in the interval [-1,1].

2.4 Neural Network Prediction Model Construction

- (1) The number of neurons in the input layer is determined

Part of the signal source is the input layer of RBF neurons. In this process, the number of signals that affect network security is determined by the number of neurons. Analyzing the security status factors, we can see that the number of neurons at the entry level is three.

- (2) The number of neurons is determined

The level switching function hidden in the neuron is a non-linear function, which is academically calculated at the center point. Adjustments can cover target differences.

- (3) Determine the number of neurons in the output layer

When using RBF to predict the network security situation, the predicted result will reflect the network security situation at the next moment, that is, a certain situation value, so the output layer only needs one node.

- (4) Selection of initial learning rate

In the training process of the RBF neural network, the size of the training rate is an important factor that determines the weight change of each neural network of the neural network, so the selection of the training rate must be accurate. If the learning rate is too high during the selection process, the entire neural network training process will become unstable. Therefore, in order to calculate the duration of the training cycle while considering system stability, the selection of learning rate is set to 0.01 to 0.7 when predicting network conditions.

- (5) Selection of expected error

Whether the network is compact and whether the selection of training samples is normal should be measured by error indicators. In general, the quadrilateral error is called the expected neural network error, and the size of the expected error should be determined according to the obvious automatic correction function.

3 Experimental Research on Network Security Design of E-Commerce Sales Management System Based on Neural Network Algorithm

3.1 Sample Settings

This article takes the e-commerce marketing management system as an example to test the effect of optimizing the network risk assessment model through automated exercises. The questionnaire method is mainly obtained by combining part of the risk assessment tool, of which 60 are used as sample combinations and 6 are used as test samples. The samples are scaled using obfuscated evaluation methods to facilitate training and testing. Due to the limitations and deficiencies of previous assessments, the original risk assessment system was combined to modify and simplify the risk assessment factors, and the results of the risk assessment were scaled to on-site values to facilitate training and network testing.

3.2 Parameter Setting and Some Parameter Research

The test receives the traditional 3D BP neural network system, the number of neurons in the input layer is $I = 10$, which is the 6 dangerous elements in the structure model, and the number of neurons in the output layer is $O = 2$, which is the risk result of the risk assessment quantity. Through experimental tests, the number of neurons in the latent nucleus is set to $H = 6$.

4 Research and Analysis of Network Security Experiment of E-Commerce Sales Management System Based on Neural Network Algorithm

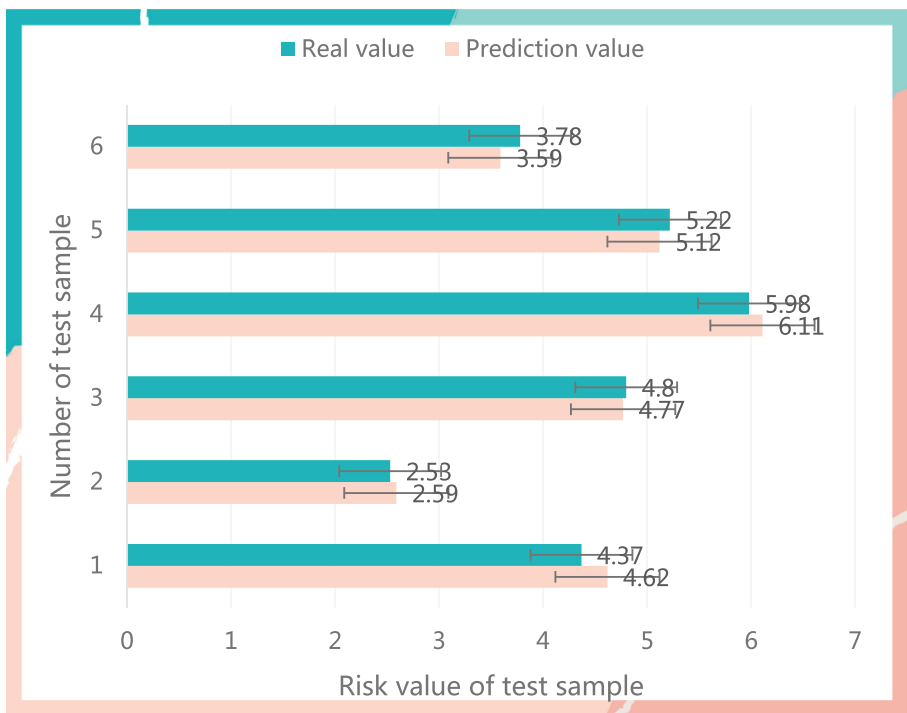
4.1 Analysis of the Effect of Neural Network Independent Experiment Detection Algorithm

In order to test the stability of the proposed method, in view of the randomness of the experimental results of the neural network, the effect of the algorithm was tested through multiple independent experiments. Table 1 shows the comparison between the average value obtained from multiple experiment results and the original risk value.

As shown in Fig. 1, it can be seen that the predicted value of the simulation experiment result is very close to the actual risk assessment result, and the error between the two is less than 2%. This shows that the algorithm can predict the results of the risk assessment.

Table 1. Comparison of experimental results and predicted results

Number of test samples	Prediction value	Real value
1	4.62	4.37
2	2.59	2.53
3	4.77	4.80
4	6.11	5.98
5	5.12	5.22
6	3.59	3.78

**Fig. 1.** Comparison of experimental results and predicted results

4.2 ICSBP Algorithm Performance Experimental Analysis

In order to further test the performance of the ICSBP algorithm, this article takes 4 algorithms as examples, and takes the average result of 6 experiments as the final result of the experiment, as shown in Table 2:

Table 2. Comparative experimental results

Sample number	Target value	BPNN	CSBPNN	ICSBPNN	GABPNN
1	4.2	4.721	4.319	4.231	4.353
2	3.5	3.574	3.421	3.478	3.456
3	2.4	2.582	2.610	2.374	2.425
4	4.8	4.893	4.826	4.713	4.883
5	5.2	5.267	5.341	5.182	5.167
6	3.7	3.681	3.706	3.716	3.694

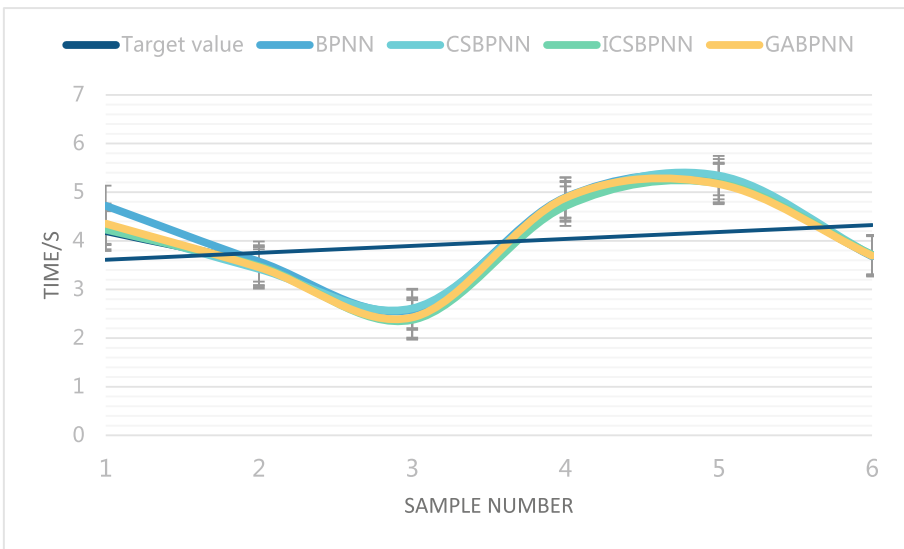


Fig. 2. Comparison of experimental results

As shown in Fig. 2, it can be seen from the experimental results that the execution time of the BP neural network is the shortest, one-fifth less than the time of other algorithms. The other three algorithms have slightly different processing times. The prognostic effect of BP neural network has large errors and cannot accurately predict the results of risk assessment.

5 Conclusions

Analyzing and predicting the state of network security is the focus of this article. Through network monitoring, obtain several factors related to network security services, analyze these factors, and use the network security evaluation model to obtain the current network performance status. At the same time, the network status prediction function. It was created based on the results of the assessment to enable network administrators to better

understand the future network development culture, respond to potential security issues in a timely manner, and maintain network security and stability.

Acknowledgements. This work was supported by Innovation and Entrepreneurship Training Program for College Students of Xi'an Fanyi University (No. 202112714005).

References

1. Liang, C., Zhang, Q., Ma, J., Li, K.: Research on neural network chaotic encryption algorithm in wireless network security communication. *EURASIP J. Wirel. Commun. Netw.* **2019**(1), 1 (2019). <https://doi.org/10.1186/s13638-019-1476-3>
2. Du, Y.A.: Research on the route pricing optimization model of the car-free carrier platform based on the BP neural network algorithm. *Complexity* **2021**(4), 1–10 (2021)
3. Soni, N., Sharma, E.K., Kapoor, A.: Hybrid meta-heuristic algorithm based deep neural network for face recognition. *J. Comput. Sci.* **51**(June), 101352 (2021)
4. Maddodi, G., Awad, A., Awad, D., Awad, M., Lee, B.: A new image encryption algorithm based on heterogeneous chaotic neural network generator and dna encoding. *Multimedia Tools Appl.* **77**(19), 24701–24725 (2018)
5. Chen, L.-P., Yin, H., Yuan, L.-G., Lopes, A.M., Machado, J.A.T., Wu, R.-C.: A novel color image encryption algorithm based on a fractional-order discrete chaotic neural network and DNA sequence operations. *Front. Inf. Technol. Electron. Eng.* **21**(6), 866–879 (2020)
6. Guo, L.: Research on anomaly detection in massive multimedia data transmission network based on improved PSO algorithm. *IEEE Access* **8**, 95368–95377 (2020)
7. Hammad, M., Zhang, S., Wang, K.: A novel two-dimensional ECG feature extraction and classification algorithm based on convolution neural network for human authentication. *Future Gener. Comput. Syst.* **101**, 180–196 (2019)
8. Liu, F., Huo, W., Han, Y., Yang, S., Li, X.: Study on network security based on PCA and BP neural network under green communication. *IEEE Access* **8**, 53733–53749 (2020)
9. Wang, X., Ma, D., Hu, K., Hu, J., Du, L.: Mapping based residual convolution neural network for non-embedding and blind image watermarking. *J. Inf. Secur. Appl.* **59**, 102820 (2021)
10. Siregar, B., Habibie, I., Nababan, E.B., Fahmi,: Identification of Indonesian clickbait news headlines with long short-term memory recurrent neural network algorithm. *J. Phys. Conf. Ser.* **1882**(1), 012129 (2021)
11. Jia, Y., Wang, M., Wang, Y.: Network intrusion detection algorithm based on deep neural network. *IET Inf. Secur.* **13**(1), 48–53 (2019)
12. Wu, L., Zhou, J., Li, Z.: Applying of GA-BP neural network in the land ecological security evaluation. *IAENG Int. J. Comput. Sci.* **47**(1), 11–18 (2020)



Data Mining and Computer Software Database Programming Technology

Xiaoqing Wang^(✉)

School of Information, Wuhan Vocational College of Software and Engineering,
Wuhan 430025, Hubei, China

Abstract. With the development of computer technology, the application provided by the database has far exceeded its original data function. It can form a knowledge base and combine with data mining technology to form a database programming system and play an important role in the development of programming languages. This article establishes a database programming model on the basis of data mining and computer software, uses JAVA, Python, PHP, C++ and other programming techniques to test the model's control effect on programming instructions and programming speed, and find a suitable database programming model for this article. Programming techniques to develop programming scripts.

Keywords: Data mining · Computer software · Database editing technology · Programming language

1 Introduction

In the context of the rapid development of information technology, the performance of data mining and computer software technology is getting stronger and stronger. The application of these technologies in various fields is gradually expanding, and it also brings unprecedented opportunities to programming. Although computer programming has begun to be gradually applied in the traditional programming process, the failure of programming cannot be avoided in the face of increasing programming data. The performance of data mining technology to quickly process massive data just meets the needs of computer programming. Therefore, the introduction of data mining technology in the programming process is of great significance to improve the programming speed.

At present, many scholars at home and abroad have discussed data mining and computer software database editing technology, and have obtained corresponding research results. For example, a researcher is committed to studying various data index measurement systems, establishing a programmable data model, providing quality assurance for data transmission, programming and operation, meeting the needs of programming configuration and performance management, and realizing remote programming operations. Multi-level management and monitoring [1]. With the development of information technology, data information has proliferated. In the face of massive information, data mining technology can be used to discover hidden data information. The code information in the programming field is even more complex and changeable. A scholar just used

the data mining function to establish The special programming model greatly improves the programming efficiency [2]. Although the research on data mining and computer software database editing technology is progressing smoothly, there are still few applications that introduce data mining and computer software technology to improve the quality of programming, and people can use technology to simplify the process.

This article introduces the programming features of JAVA dynamic programming technology, which is the most commonly used programming technology. It has super computing power in the process of executing code operations. It builds a database programming model based on data mining technology and computer software, and compares JAVA technology with the other three. This kind of programming technology is applied to the programming model to control the effect of programming instructions and test the operation speed of these programming technologies, which can provide reference for the computer software database model to select the best programming technology to improve programming efficiency.

2 Java Technology and Database Programming

2.1 Java Virtual Machine Dynamic Programming Technology Characteristics

(1) Make the application hardware-independent

Virtual machines have different versions of implementations in different software and hardware environments, and each implementation provides a unified interface, follows standard specifications, and collects differences in software and hardware. JAVA code only needs to be developed once and can be deployed in different scenarios. Any legal JAVA code can run on any standard JAVA virtual machine, regardless of the difference between the underlying operating system and hardware. This allows JAVA code to be developed in accordance with a unified specification, and a single development has the ability to cross-platform and software and hardware compatibility. JAVA virtual machine operates through the stack, reducing the dependence on registers [3].

(2) Complete control over the execution of the code

All codes executed in the virtual environment are strictly monitored and programmed. The virtual machine opens up various storage spaces internally, which contains an overview of the loaded classes and what is being created, as well as the thread schedule that runs on the stack. In the execution process, it has a complete understanding of the structure and distribution of the space and the characteristics of each loading type, controls the storage locations and index references of all programming objects, manages all storage spaces, and fully understands the current settings of each thread Implementation status and conditions. In short, the virtual machine controls the execution of all codes, and all operations are in a controlled state. This is what traditional applications lack. Due to these characteristics, the virtual machine can perform auxiliary editing or adjust the code mode in real time during the normal code interval [4].

(3) Have strong operation ability

The virtual machine reports some internal information through related interfaces and provides a variety of external control methods, enabling us to analyze and control application execution from the virtual machine level. It can span all types of loads, modify its properties and control code execution. The virtual machine provides us with a very effective way to analyze the runtime behavior of the code and control its operation. Although this operation capability is limited by the functions provided by the interface, theoretically we can modify the code of the virtual machine to achieve a wider range of purposes [5].

(4) Flexible execution method

The code execution process can be understood as execution or compilation or a mixed execution of the two, depending on the setting of the code. The advantage of compilation and execution is high performance, but the compilation process takes a long time. The advantage of interpreted execution is that the code form can be modified during execution. This feature is very attractive to developers of Basic programs, because the code can be modified arbitrarily during the debugging process and can continue to execute without restarting. The JAVA virtual machine has the ability to dynamically switch between the two operating modes. This allows us to take into account performance and flexibility, and allows us to dynamically modify the code while it is running. JVM is used as a runtime execution environment. Every Java program in operation is actually run in an instance of JVM, and each running program is run by its own JVM instance [6].

2.2 Classification of Database Programming

The method of database programming can usually be divided into manual programming and automatic programming. Manual programming refers to the programming process of all steps from partial mode analysis, data calculation, batch program collection to program verification. These steps are all completed manually. It is suitable for relatively simple calculations, not many programming parts, and uncomplicated programming. Numerical calculations are quite complex, and those that are difficult to correct due to large calculation workloads become programs that are difficult to achieve through manual programming.

Automatic programming, also known as computer programming, refers to the process in which a computer automatically performs numerical calculations and executes compilation. Program developers only need to compile the requirements, write a relatively short code in the designated scripting language, and then enter the computer, and the computer will automatically execute the code programming, so that the programmer can check whether the program is correct in time and make necessary Modify to get the correct programming instructions. Automated programming replaces programmers to perform tedious arithmetic operations, reduces their workload, and improves programming accuracy by reducing human calculation errors. Automatic programming is accomplished through computer-aided numerical programming technology, and special database programming software needs to be established [7, 8].

2.3 Development Characteristics of Database Programming Technology

(1) Integration of database programming system

Integration refers to the integration of database programming into the computer software system. At present, in computer software systems, two integration methods of communication connection type and database type are widely used. Commonly used in the communication connection mode is to establish a standard data communication mode to connect the database programming system with the computer software system. Database integration refers to the unified management of various data in the process of cleaning, conversion, and management by creating a public database, and all programs are only connected to the database. Therefore, the database is not only a public information source, but also a public access point. At present, the most widely used database integration method is the compact geometric modeling database integration method. This method uses the human-computer interaction function to extract the necessary geometric and topological information from the database for database programming.

(2) Intelligentization of database programming system

Intelligentization refers to the introduction of human thinking into an integrated computer software system, and the allocation of human thinking judgment and decision-making methods to machines for completion. Therefore, it is necessary to use artificial intelligence methods to create knowledge bases of different systems and change the role of human decision-making in the process of solving various problems. In addition, there is an integrated measurement, programming and processing system, which is specially used to test a certain program, and then enter the database through the access port for automatic control and programming [9].

(3) Modularization and generalization of post-processing programs

The post-processing program should convert the calculated relevant data information into the input information of the specific database according to the different databases used. A recently developed “modular post-processing program” is to integrate some common functions of different databases into a standard model. When programmers need a new post-processing program, all they have to do is to develop a “driver”, and just select some standard modules to add to the database [10, 11].

2.4 Auxiliary Data Mining Based on Data Reasoning

When a new program appears, first describe the content of the program according to the specified method, then retrieve similar programs from the database according to the data mining algorithm, and again use the database closest to the new program as its data storage; then according to the knowledge base and the characteristics of the new program Load the memory to make it more suitable for the new program to run, and finally store the new program as new data in the database for use [12].

$$\text{sim}(a_i, b_i) = 1 - \frac{2 \times |a_i - b_i|}{|a_i + b_i|} \quad (1)$$

Among them, a_i and b_i ($i = 1, 2, \dots, n$) are numerical attributes, respectively, and $sim(a_i, b_i)$ is the similarity between attributes.

$$sim(c_1, d_1) = \frac{\sum_{i=1}^n w_i^{sim}(a_i, b_i)}{\sum_{i=1}^n w_i} \tag{2}$$

Among them, c_1 and d_1 are two data in the database, and their similarity is represented by $sim(c_1, d_1)$, and w_i is the weight of the i -th attribute.

3 Based on Data Mining Technology and Computer Software Database Programming Model

3.1 Function Module

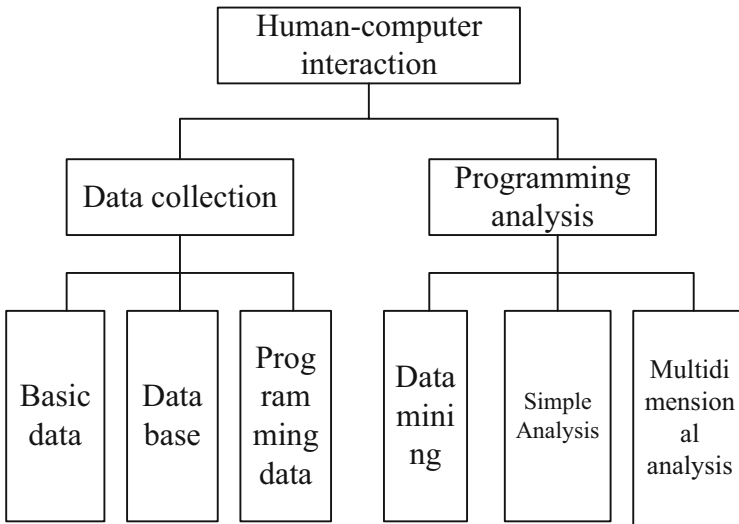


Fig. 1. Computer software database programming function module

This article introduces data mining into database programming, and establishes a computer programming model based on database and data mining. According to the needs of programmers, set the system function module as shown in Fig. 1. The model can obtain programming-related data and information, and clean up and convert them to eliminate irregular data; it can store programming data according to different programming themes, and display it in front of programmers in a multi-dimensional structure; According to different programming requirements, different programming methods are selected for programming analysis to improve the accuracy of programming results.

3.2 The Overall Structure of the Programming Model Based on Data Mining Technology

As shown in Fig. 2, it is based on data mining technology and computer software database programming model. Among them, the database cleans, converts, and loads the information extracted by the information collection system, and stores the processed data according to the programming theme. At the same time, the database presents these processed data to programmers in a multi-angle and multi-level manner according to the requirements of programmers. The data stored in the database is integrated, consistent, and high-quality data, which brings a lot of convenience to the subsequent programming work.

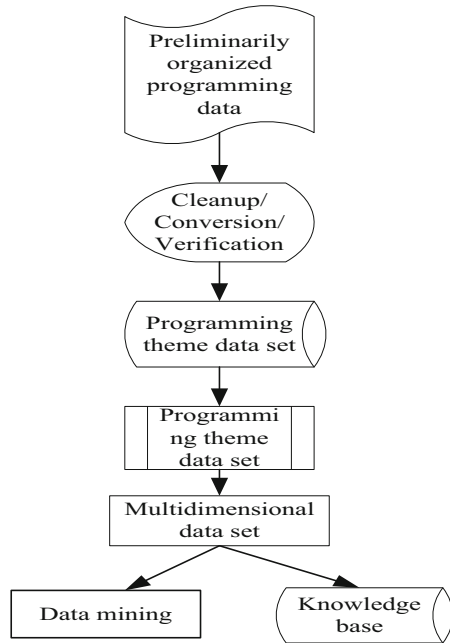


Fig. 2. The overall structure of the programming model

4 Performance Test of Database Programming Model

4.1 Automated Testing

The tools for automated testing in this article are JAVA, Python, PHP, C++. Use these programming techniques to develop programming languages and test the automated performance of the database programming model of this article under each technology. Based on CLI (command line program), API (application programming interface) and GUI (graphical user interface) testing, that is, the application of scripting technology to

Table 1. Control effects of different programming techniques on programming instructions

	CLI	API	GUI
JAVA	78%	80%	86%
Python	85%	75%	81%
PHP	77%	84%	82%
C++	89%	82%	83%

send CLI commands, API requests or GUI requests to the device simulation to achieve the effect of controlling the device. The test results are shown in Table 1. The programming model uses Java technology to have the highest control effect on GLI commands, Python technology has the best control effect on CUI requests, PHP technology has the best control effect on API requests, and C++ technology has the best control effect on CLI effects on programming instructions, and each has its own advantages.

4.2 Programming Technology Speed Test

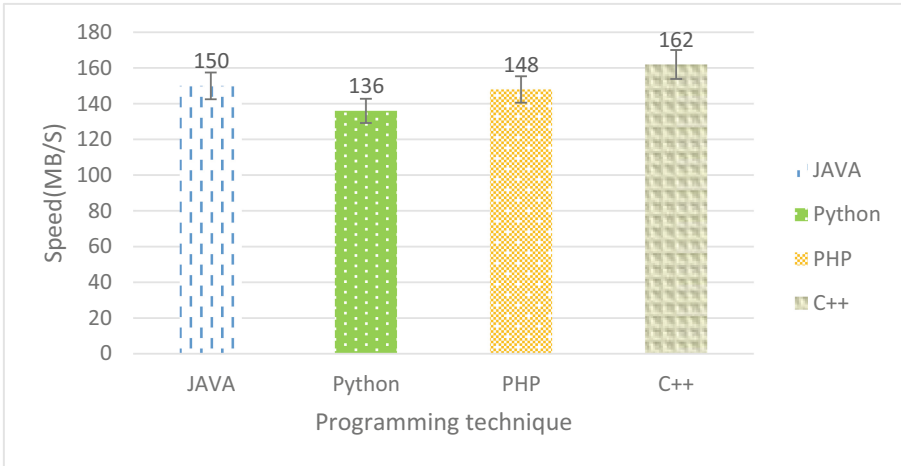


Fig. 3. Programming speed of different programming techniques

As shown in Fig. 3, different programming technologies are used in the database programming model to develop programming scripts to test the programming speed of the model under different programming technologies. The result is that the programming speed using Java technology is 150 MB/s. The programming speed using Python technology is 136 MB/s, the programming speed using PHP technology is 148 MB/s, and the programming speed using C++ technology is 162 MB/s. It can be seen that this

model is more suitable for programming with C++ preview, as the amount of programming when large, it can effectively improve programming efficiency compared to other programming techniques.

5 Conclusion

This paper establishes a database programming model based on data mining technology and computer software, and conducts research and analysis on the programming instruction control effect and programming speed of different programming techniques used in the programming model. It is found that in terms of programming instruction control, each technology is in CLI Command control, API request control and CUI request control have their own advantages and disadvantages. In terms of programming speed, the programming speed of C++ technology in the programming model is significantly higher than other programming technologies.

References

1. Brusilovsky, P., Malmi, L., Hosseini, R., Guerra, J., Sirkiä, T., Pollari-Malmi, K.: An integrated practice system for learning programming in Python: design and evaluation. *Res. Pract. Technol. Enhanc. Learn.* **13**(1), 1–40 (2018)
2. Clement, T.E.: Coding literacy: how computer programming is changing writing. *J. Am. Soc. Inform. Sci. Technol.* **69**(9), 1174–1176 (2018)
3. Cardenas-Cobo, J., Puris, A., Novoa-Hernandez, P., Galindo, J.A., Benavides, D.: Recommender systems and scratch: an integrated approach for enhancing computer programming learning. *IEEE Trans. Learn. Technol.* **13**(2), 387–403 (2020)
4. Rakhimova, A.E., Yashina, M.E., Mukhamadiarova, A.F., Sharipova, A.V.: The development of sociocultural competence with the help of computer technology. *Interchange* **48**(1), 55–70 (2016)
5. Chow, J.C.L.: Internet-based computer technology on radiotherapy. *Rep. Pract. Oncol. Radiother.* **22**(6), 455–462 (2017)
6. Wang, X.M., Hwang, G.J., Liang, Z.Y., et al.: Enhancing students' computer programming performances, critical thinking awareness and attitudes towards programming: An online peerassessment attempt. *Educ. Technol. Soc.* **20**(4), 58–68 (2017)
7. Sun, Q., Wu, J., Rong, W., et al.: Formative assessment of programming language learning based on peer code review: implementation and experience report. *Tsinghua Sci. Technol.* **24**(04), 55–66 (2019)
8. Díaz, J., López, J.A., Sepúlveda, S., et al.: Evaluating aspects of usability in video game-based programming learning platforms. *Procedia Comput. Sci.* **181**(2), 247–254 (2021)
9. Unal, A., Topu, F.B.: Effects of teaching a computer programming language via hybrid interface on anxiety, cognitive load level and achievement of high school students. *Educ. Inf. Technol.* **26**(5), 5291–5309 (2021)
10. Ivanovic, M., Xinogalos, S., Pitner, T., et al.: Technology enhanced learning in programming courses – international perspective. *Educ. Inf. Technol.* **22**(6), 2981–3003 (2017)
11. Desai, S., Desai, S.: The scheme programming language. *Int. J. Comput. Sci. Eng. Inf. Technol. Res.* **9**(1), 13–22 (2019)
12. Gennaro, D.L., Li, Z., Sami, M., et al.: Visual programming language environment for different IoT and robotics platforms in computer science education. *CAAI Trans. Intell. Technol.* **3**(2), 119–130 (2018)



Design of Visualization System Under Centralized Monitoring of Regional Power Grid Based on Data Mining Algorithm

Piao Liu, Jinman Lou, Kai Yu, Xiaoxia Li^(✉), and Youjun Feng

Dongguan Power Supply Bureau of Guangdong Power Grid Corporation,
Dongguan 523000, China

Abstract. With the advancement of the construction of the State Grid project, intelligent power supply has become a necessary condition for improving the quality of power supply of the power grid. At the same time, with the increasing level of intelligent power supply management, more monitoring devices have been adopted. The number of these monitoring devices is huge, and effective data collection and data processing must be realized through the supporting monitoring management system. Therefore, the establishment of an intelligent power grid monitoring system is a necessary and critical issue. This paper studies the visualization system under the centralized monitoring of the power grid, and understands the related theories of the visualization system and data mining under the centralized monitoring of the power grid on the basis of literature data, and then designs the visualization system under the centralized monitoring of the regional power grid based on the data mining algorithm. And then detect the mining algorithm used by the system. The result of the test is that the accuracy of the algorithm cited in this article is about 86%, which is higher than that of the traditional algorithm.

Keywords: Data mining · Visualization system · Power system · Smart grid

1 Introduction

The power grid has a wide distribution range and contains a large amount of equipment data [1, 2]. Related power distribution equipment needs to be monitored and remotely controlled, such as real-time understanding of the actual operation of the distribution network, and other advanced applications such as distribution network, power flow analysis, line loss management, etc. [3, 4]. Due to the rapid changes in the operation of the power system, it is difficult for operation and maintenance personnel to deal with emergencies such as network failures and disasters. The demand for distribution network automation has given birth to the emergence of distribution network monitoring systems, distribution network maintenance, function research, software design, and automation monitoring system hardware [5, 6]. The power grid is a collection of rapid changes, large specifications, and complex structure. Its main function is a grid automation monitoring system at the level of grid operation, maintenance and management, but its automation needs to be further improved [7, 8].

Regarding the research of power grid monitoring, some researchers pointed out that substations play a central role in the transmission and distribution of modern electric energy interconnection, and its safe operation is particularly important. After the establishment of the substation, it is necessary to solve the problems of safe and reliable operation of substation equipment, safety management of substation personnel, and remote monitoring. The substation has a monitoring system with “four remote” functions, but there is still a lack of specific solutions for how to comprehensively monitor the working environment of the equipment and the safety status of the entire substation. It is useless only in the regular monitoring device for shift workers at the central monitoring core, because the on-duty personnel cannot monitor the entire electrical equipment and the surrounding environment 24 h a day. Even if a video monitoring system is installed, the observation effect can be timely. It is very troublesome to find out the emergencies correctly, because all kinds of traffic accidents happen in blind spots. Therefore, it is necessary to establish a safe and stable working environment for various factors that affect the operating conditions of all facilities in the substation and the safe operation of the substation [9]. Relevant research and development personnel believe that the current monitoring and control system is facing the problem that, in addition to the traditional “four remote” functions, it is necessary to monitor the appearance of the substation facilities, monitor the environment such as fire protection and anti-theft of the substation, and control the change the working environment of the power station, such as fire prevention and theft, is monitored. Since its main purpose is to prevent fire, blasting, leakage, theft, malicious damage to equipment, etc. that cause or cause serious danger to the safe environment, the monitoring scope mainly involves the substation environment and the original four remote controls. The current “five remote” technology can be said to be China’s first remote video surveillance technology implemented in banks, road transportation, universities, customs, business chains and other places, and mature local surveillance technology in unmanned substations. At present, there is a relatively complete monitoring system solution, but it is still not well connected with the basic management of the power industry, especially the monitoring of changes in the status of primary equipment, and accurate monitoring of substation equipment [10]. To sum up, there are many research results on power grid monitoring, but the research on the design of monitoring system needs to be studied in depth.

This paper studies the visualization system under the centralized monitoring of the power grid, analyzes the requirements of the visualization system under the centralized monitoring of the power grid and the application of data mining on the basis of literature data, and then designs the visualization system under the centralized monitoring of the regional power grid based on the data mining algorithm. And then the system is tested, and draw relevant conclusions through the test results.

2 Research on Visualization System and Data Mining Under Centralized Monitoring of Power Grid

2.1 Visual System Requirements Under Centralized Monitoring of the Power Grid

- (1) Multiple detection and intrusion detection functions. Different substation locations require different detection equipment to detect and detect intrusions. Different detectors are organically combined to form front-end detection equipment, forming a safety net for unattended substations.
- (2) Intrusion video connection monitoring alarm function. When the system detects a rogue intrusion signal, the camera system installed in the substation can monitor the alarm location in time, and perform video surveillance and recording.
- (3) Video connection monitoring and alarm function when equipment failures such as substations and environmental abnormalities occur. Once the system detects the fault of the substation equipment or the abnormality of the surrounding environment of the equipment, the camera system can timely monitor the faulty equipment or the abnormality of the environment, thus realizing video monitoring and camera.
- (4) Alarm sound and light alarm function. When the system detects illegal entry into the substation, the control system can activate the sound and light alarm to generate high-frequency alarms.
- (5) Remote video and alarm signal transmission function. The system can alarm for illegal intrusion, substation facility failure, environmental failure and other alarms, and generate alarm signals, which are then sent to the remote transmission monitoring center through the wireless network. At the same time, the remote monitoring center can also collect the video monitoring signals of the substation in real time.

2.2 Data Mining Applications

- (1) Correlation analysis

All data stored in the database have a specific relationship or a specific relationship between the values of many variables. The purpose of data correlation analysis is to reveal these hidden correlations in the database [11].

The rules generated by association analysis are reliable because it is impossible to determine whether there is an association function between the data stored in the database. By extracting the association rules derived from the data, you can further understand customer behavior and so on. The algorithm used in the correlation analysis mainly includes the Aprii algorithm.

- (2) Cluster analysis

Unlike classification, the class label of each object in the cluster is unknown, and grouping is the process of dividing data objects into classes or clusters. The objects in the cluster are similar to each other, but completely different from the objects in other clusters [12]. Cluster analysis is generally considered because the data does not have

a certain amount of descriptive information and cannot be organized into classification standards.

2.3 Data Mining Algorithm

Data mining algorithms seem very simple, but the actual application of many complex algorithms is not satisfactory. The main functions are as follows.

- 1) Enter the “Euclidean distance”. Use this function to calculate the distance between points, set $A = \{x_1, x_2, \dots, x_m\}$, and set the sample Euclidean distance to d_{ij} , do the following:

$$d_{ij} = \sqrt{(x_{i1} - x_{j1}) + (x_{i2} - x_{j2}) + \dots + (x_{in} - x_{jn})} \quad (1)$$

- 2) A square error function used to judge whether the difference between the new cluster center and the original cluster center converges. Let x be the average of all data in cluster c_k , and c be the center of the cluster.

$$J_c = \sum_{i=1}^k \sum_{x \in c_k} |c - x|^2 \quad (2)$$

3 Design of Visualization System Under Centralized Monitoring of Regional Power Grid Based on Data Mining Algorithm

3.1 System Design Goals

The setting and management functions of Toseki power supply station are all integrated into the system, and a remote video monitoring and control system is set up in the dispatching center to remotely implement the on-site video monitoring, equipment operation monitoring and peripheral security of the substation. This is a new scientific calculation and visual shipping operation, can greatly improve the operation safety of substation, easier to repair. The topic of this paper is the construction and improvement of substation video monitoring system for the special geological conditions of the region, taking the substation as the base station of “unsupervised substation”, specially designed a visual monitoring system. Its general purpose is to:

Real-time surveillance: Set up high-resolution cameras near transformers, critical equipment and switches. The center of the power supply bureau is responsible for monitoring and scheduling the transformer and its surrounding safety, the operation of electrical equipment and remote switching operations under management. The video monitoring platform accurately presents the spatial area and image of the alarm. Currently, Toseki has 19 substations in total, and the control system design can finally present the location map of each substation, the electronic primary equipment plan, the perimeter safety diagram, and the following construction objectives.

Rapid response: combined with built-in substation automation system, achieve rapid alarm, rapid installation and access to relevant video alarm, real-time scene display,

command ordered, passive to active, video alarm can realize more and more functions such as improving the safety of the electrical system To prevent false alarm system and staff under the bad environment caused by the on-site inspection of failure.

Comprehensive application: provide the interconnection between video surveillance and comprehensive automation system, build a complete monitoring coordination control platform, meet the needs of unattended substation.

3.2 Implementation Strategies for Monitoring Data Collection

According to the analysis of the above concepts, the monitoring data acquisition task is divided into multiple data acquisition units for distributed parallel processing. During the processing, the synchronous clock generator is monitored, and each data acquisition unit processes the collected data, and then it records the time stamp of the data sent to the synchronous clock generator for processing, and then transmits it. The details are as follows.

1) Choose a distributed parallel processing method

This article uses the symmetric model (SPMD). The reasons are as follows:

The multi-program mode (MPMD) is not used because the work of the collection unit is equal. This helps to efficiently distribute large amounts of data and achieve load balancing. In addition, the corresponding program tasks can be configured and managed in a single way to solve data sharing and complex calculation problems, instead of using the master-slave feature. This is to prevent excessive pressure on the master process in a highly synchronized operating environment, resulting in the main process is congested, reducing the efficiency of data collection.

2) Solve the problem in a symmetrical mode

Under the symmetrical model, it is necessary to deal with the limitation of the number of parallel programs and the problems of data processing and data transmission. Once the corresponding solutions cannot be taken, the system will compete for resources or even deadlock. Therefore, this article will solve it through the badgering algorithm, and establish each process as a logical ring of a distributed system to symbolize the ability of the system, thus becoming the benchmark of whether a process can enter the critical zone. To achieve mutual exclusion of programs, you can specify access rights by setting a token on the operating system. The badge itself is also a special form of message, usually one byte long, and it keeps circulating in the logical loop of the process. Each step in the loop has a precursor and a meta report. When the token in the loop is received by a specific process through the loop, assuming that a certain process needs to enter the critical section, then the process owns the token and enters the critical section. When being pushed out of the critical section, the badge is sent to the follow-up process. Assuming that the process that has accepted the token does not need to enter the critical area, the token can be passed directly to the following steps. But because the logic circle is just a badge, the processes are mutually exclusive.

3.3 Visual Display System Data Source

From the perspective of different data applications, data belongs to different categories, and different data types (data types shown in Fig. 1) correspond to different storage, calculation, and transmission requirements and technical implementations.

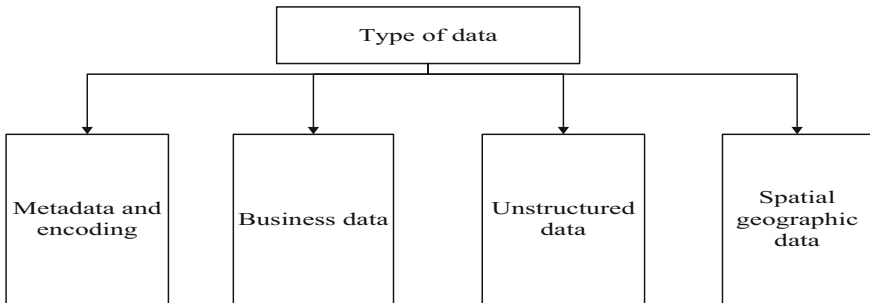


Fig. 1. Visual display system data source

1) Metadata and coding

Detailed business data model SG-CIM: defines an integrated business data model covering core business systems.

Index model: The definition of a comprehensive index model.

Analysis model: The definition of a comprehensive analysis model.

Standard coding: a unified data coding standard.

2) Business data

Company source details: China State Grid's business application system business detailed data. Use the data generated by each business data model.

Business detail data management: From the business detail data of the application management system of various industries in China.com, the data format is transformed into a new data type that meets the coding standard.

Indicator data: Through big data analysis, the summary data of different monitoring indicators is exported to different business systems and corresponding auxiliary systems.

3) Unstructured data

All kinds of unstructured data captured in operation monitoring analysis, including various reports and reports automatically generated after data processing, are specifically used to support operation monitoring and analysis, as well as data, rules, systems, and regulations. This document contains the following items: the first category of declarations, documents, reports, regulations, systems and abstracts, the second category of tracking statements, reports, the third category of analytical statements, reports, and the

fourth category of coordinated management documents. Including various video, audio and other multimedia data for monitoring and analyzing features.

4) Spatial geographic data

Spatial geographic information can be displayed using various map data required for analysis based on GIS systems, including basic geographic data such as cities, regions, and roads.

4 System Test

4.1 Data Source

The experimental data is taken from the electricity consumption data of 1 million households in a specific power grid, with a sampling frequency of 10 min/h and a sampling time of 24 h, with a total of 144 dimensions. Each experimental data source contains only 144 power consumption points. It represents 20 bytes. At the same time, it also collects the amount of electricity consumption information, electrical appliances, and households in the home area of these household users for experimental reference.

4.2 Experimental Design

The purpose of this experiment is to compare the accuracy of optimized and non-optimized algorithms. From the experimental data sources, common low-power, medium-power and high-power household users are randomly selected as the test source data. Each data set contains three types of data: high power consumption, medium power consumption and low power consumption. It uses optimized and non-optimized algorithms to process the aforementioned data configured to run on the same independent computer. The initial grouping value of the non-optimized algorithm is set to 3. In the whole experiment process, there are also 3 optimization algorithms to avoid some random factors, repeat the same experiment 10 times, and finally take the average value. The experimental results are shown in Table 1.

Table 1. Accuracy of optimized and non-optimized algorithms

	Not optimized correct rate/%	Correct rate after optimization/%
4000	79.4	87.7
8000	78.0	89.5
12000	71.3	86.3
16000	68.1	88.9
20000	66.0	87.2

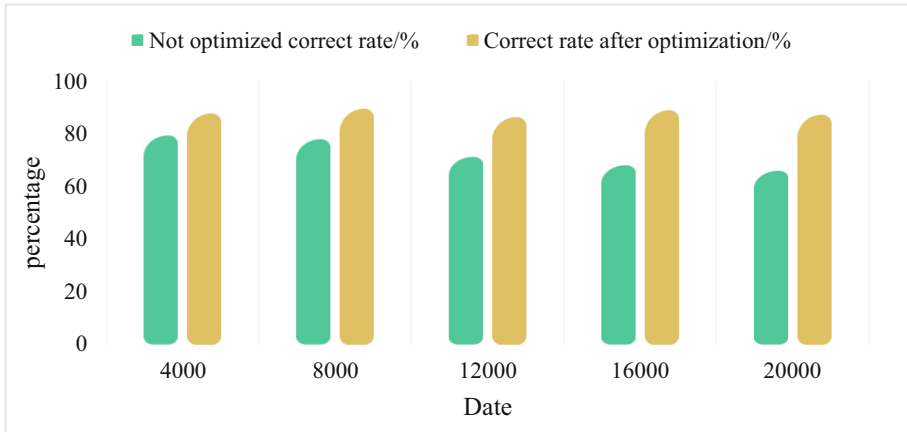


Fig. 2. Accuracy of optimized and non-optimized algorithms

It can be seen from Fig. 2 that as the amount of data increases, the accuracy of the non-optimized algorithm decreases. After preprocessing, the initial selection of the grouping center is performed correctly, and the accuracy is relatively constant and remains at about 87%. The optimized algorithm can be derived from experiments, and it is more accurate than the non-optimized algorithm.

5 Conclusions

This paper studies the visualization system under the centralized monitoring of power grids. After understanding the relevant theories, design the visualization system under the centralized monitoring of regional power grids based on data mining algorithms, and test the data mining algorithms used in the system. The test results It is concluded that the optimized algorithm in this paper has a higher accuracy rate than the unoptimized algorithm.

Acknowledgements. Key Informatization Projects of China Southern Power Grid Co., Ltd. (031900HK42200008).

References

1. Inoue, A., Ikeda, Y., Yatabe, K., et al.: Three-dimensional sound-field visualization system using head mounted display and stereo camera. *J. Acoust. Soc. Am.* **140**(4), 3195–3196 (2016)
2. Grudzien, K.: Visualization system for large-scale silo flow monitoring based on ECT technique. *IEEE Sens. J.* **17**(24), 8242–8250 (2017)
3. Oca, I.M.D., et al.: Accuracy of toric intraocular lens axis alignment using a 3-dimensional computer-guided visualization system. *J. Cataract Refractive Surg.* **42**(4), 550–555 (2016)
4. Yusof, R., Qazi, A., Inayat, I.: Real-time student visualization system in classroom using RFID based on UTAUT model. *Campus-Wide Inf. Syst.* **34**(3), 274–288 (2017)

5. Velázquez-Iturbide, J.Á., Pérez-Carrasco, A.: How to use the SRec visualization system in programming and algorithm courses. *ACM Inroads* **7**(3), 42–49 (2016)
6. Kawamura, Y., Moriyama, Y., Jang, H.: Web-GIS based visualization system of predicted ground vibration induced by blasting in urban quarry sites. *J. Geogr. Inf. Syst.* **11**(1), 17–31 (2019)
7. Kwon, K.-C., et al.: Advanced three-dimensional visualization system for an integral imaging microscope using a fully convolutional depth estimation network. *IEEE Photonics J.* **12**(4), 1–14 (2020)
8. Ryabinin, K.V., Kolesnik, M.A.: Automated creation of cyber-physical museum exhibits using a scientific visualization system on a chip. *Program. Comput. Softw.* **47**(3), 161–166 (2021)
9. Kitahara, H., Hons, J., Thourani, V.H.: Commentary: technology to make the invisible “visible”: Novel optical visualization system for off-pump artificial chordae implantation – *ScienceDirect. J. Thorac. Cardiovasc. Surg.* **158**(5), 1341–1342 (2019)
10. Mezentsev, A., Pavlov, V., Golobokov, Y., Kabrysheva, O., Deriglazov, A.: The efficiency indicator for information visualization system of the Tokamak experimental device operating in repetitively pulsed mode. *Fusion Eng. Des.* **143**, 115–120 (2019)
11. Inazumi, S., Adachi, Y., Kizuki, T., Tsuchiya, J., Funahashi, M.: Development of construction information visualization system and estimation of N value by current value measurement on ground-improvement work. *Constr. Rob.* **3**(1–4), 103–116 (2019)
12. Smith, N., Nimbarte, A.D., Gopalakrishnan, B., Wuest, T.: Real-time energy visualization system for light commercial businesses. *Sustainable Energy Technol. Assess.* **34**, 68–76 (2019)



Application of Data Mining Algorithm in Realizing Monitoring and Visualization of Transmission Network Government Affairs

Fangtao Wan¹(✉), Shengxiang Xie¹, Jianshu Yuan¹, and Lixia Zheng²

¹ Dongguan Power Supply Bureau of Guangdong Power Grid Corporation,
Dongguan 523000, China

² Dongguan Guanneng Green Energy Service Co., Ltd, Dongguan 523000, China

Abstract. To comply with the new development trend of automation, informatization, and intelligence in the power industry, to ensure the economic operation, safety, reliability, and effective operation of the entire power system, to further adapt to the requirements of power supply companies, to further optimize the use of corporate resources and overall management, and to actively carry out the power grid. The development and practice of visual management strategies for government affairs monitoring has great practical significance. This paper aims to study the application of data mining algorithms in the implementation of monitoring visualization of transmission grid government affairs. Based on the analysis of the data mining process, the application of data mining algorithms and visualization techniques in power grid government affairs, a visualization system for grid government affairs monitoring is designed. The performance of the system was tested. The test results show that the maximum response time of the power grid government monitoring visualization system does not exceed 3s, and the system CPU occupancy rate and memory resource occupancy rate do not exceed 20%, so the performance test of the system passed.

Keywords: Data mining · Transmission grid government affairs · Monitoring visualization · Visualization system

1 Introduction

Data mining refers to a way to obtain practical information and knowledge from big data analysis. It is an industry that has emerged in recent years. Data mining can not only find a large amount of existing information and knowledge in a large amount of random, fuzzy, noisy and incomplete data, but also find a large amount of unknown and potentially useful information and knowledge, and the obtained information and knowledge are more convenient to apply and understand, Easy to store and use [1, 2].

In recent years, data mining technology has achieved many results in power grid applications, which are mainly reflected in state evaluation, user feature extraction, error diagnosis, and load forecasting. In state assessment research, many researchers use decision tree methods, neural network methods, or a combination of both. Some

researchers have proposed the application of the method based on the combination of neural network and decision tree in the electronic security assessment of power system [3, 4]. Many researchers use cluster analysis to extract user characteristics in their research, but some also use methods such as decision trees. For example, foreign researchers propose to use adaptive decision tree methods to determine attributes such as climate, temperature and humidity, and user electricity records, and correlate user attributes with other attributes for classification analysis [5, 6]. Having said that, there are more and more researches on data mining technology for power system fault diagnosis and power load application. Cluster analysis, set theory, decision tree, neural network and other methods are used in power system fault diagnosis and load forecasting research. Some researchers have discussed the use of cluster analysis to analyze the gas composition of the transformer, and obtain the relationship between the fault and the gas composition of the transformer [7, 8]. Some researchers combined with rough set theory, proposed to establish a decision table based on the fault status and monitoring parameters of power equipment, and derive the rules of equipment fault diagnosis [9, 10]. Generally speaking, various data mining techniques can be applied to power systems to eliminate noise and other data interference and find useful information, which greatly improves the efficiency of power managers' decision-making. However, some advanced data mining theoretical methods are used in power systems. The application in the system has not been fully promoted and has not really played its role [11, 12]. Therefore, we need to integrate data mining technology and various business information management systems into the power system according to the actual situation, and effectively combine them to assist manual management and scientific decision-making.

On the basis of consulting a large number of relevant references, combining the process of data mining, data mining algorithms and the application of visualization technology in power grid government affairs, this paper designs a power grid government affairs monitoring visualization system. The system mainly includes three modules, respectively It is a login module, a data conversion aggregation module and a comprehensive query module. After implementing these three modules one by one, the performance of the system is tested to verify whether the performance of the system meets the requirements of this article.

2 Application of Data Mining Algorithm in Realizing Monitoring and Visualization of Transmission Network Government Affairs

2.1 The Process of Data Mining

The most important function of the data mining process is looping. Each step has two meanings: pass or fail. If the execution is successful, go to the next step, otherwise return to the previous step. Generally speaking, the data mining process includes four main steps:

- (1) The definition of data mining problems. This step is used to determine the direction of the data mining operation. First, understand the business process according to the needs of users, roughly determine the data range, its characteristics, and relationships related to the data mining theme, and then analyze the target data. That is,

a detailed understanding of the data. The conversion from business type to theme application type determines the data dimension and data storage perspective.

- (2) Design the data model. The massive amount of original data faced by data mining is usually stored in various business databases, most of which exist in the form of business processes, so data mining cannot be performed directly. The design of the data model is based on the subject of application analysis, transforming business data into a warehouse.
- (3) Establish a mining model. Modeling is the basic step of data mining and may need to be repeated. The most logical and effective model is determined by constantly changing parameters during the modeling process. The standard modeling process is to efficiently extract target data from the data source based on the results of thematic analysis, and use appropriate data mining algorithms to create a data model. Data mining models can collect and analyze target data, and predict future growth trends based on data from previous periods.
- (4) Announce the results. This step is mainly user-oriented and needs to be improved for ease of use, because it is the most intuitive experience provided to users by the data mining system. The result version is to visually display the results of data mining to users and provide users with additional data mining capabilities.

2.2 Data Mining Algorithm

(1) ID3 algorithm

The idea of the D3 algorithm is to do more with fewer tools. Therefore, this shows that the smaller the size of the decision tree, the better, and the more accurate the prediction of the classification result. Of course, the smallest tree structure is not always created. This is just a heuristic algorithm. Some important concepts of the algorithm are:

1) Information entropy

Entropy is an indicator of the uniformity of spatial energy distribution. Information entropy is the expected value of a random variable, and is the most commonly used indicator to measure the purity of a set of samples. The greater the information entropy, the harder it is for us to draw our own conclusions. The process of establishing classification rules is essentially a process of reducing information entropy. Now we have a data set T, which contains the number of samples in the xth category and the ratio of the set. The total number of top samples is p_x ($x = 1, 2, \dots, |z|$) and the information entropy of the set T data is calculated as formula (1):

$$Entropy(T) = - \sum_{x=1}^{|z|} p_x \log_2^{p_x} \tag{1}$$

2) Information gain

Since the non-numeric variable r can take N values $\{r_1, r_2, \dots, r_m\}$, use specific variables to divide the data, and the result of the division contains n branches, where T^n represents the new branch sample series. Therefore, the “information gain” obtained by dividing the sample T by the variable r is shown in Eq. (2):

$$Gain(T, r) = Entropy(T) - \sum_{n=1}^N \frac{|T^n|}{n} Entropy(T^n) \tag{2}$$

(2) Naive Bayes Classification

As the theoretical basis of Bayesian classification algorithm, Bayes theorem occupies an important position in Bayesian algorithm. For example, given $P(A|B)$, find the probability $P(B|A)$ after A and B are exchanged. The basic equation for solving $P(A|B)$ is shown in Eq. (3):

$$P(A/B) = P(AB)/P(B) \quad (3)$$

2.3 Application of Visualization Technology in Power Grid Government Affairs

(1) Self-service data modeling

The traditional data model based on data warehouse allows users to only use the high-quality data warehouse that has been made, but through data visualization technology, users have been able to easily analyze the data in the database table. The model visualization technology establishes the model and describe the virtual table setting association in SQL language, establish the association modeling between reports, data tables and data views, and use database views as virtual tables to model data. At the same time, the data modeling tools provided by the visualization technology are independent of the customer's current data processing environment, and can also establish a completely customized data mining model. In addition, a security mechanism is proposed to allow users to browse their trusted data models freely and to communicate with others through permission. This also ensures free and open sharing of data access. And this more flexible mixed data model solution can not only implement star and snowflake models, but also create and use custom data models.

(2) Visual arrangement

The power grid government affairs monitoring visualization platform is designed with an IDE developed using pure JS code language. It can be designed to visualize data in most web pages. Designers can also set the interface suitable for mobile display devices, and provide a resolution adjustment function to achieve layout. The interface can be customized according to various display devices with different resolutions in a certain area. The visualization data provided on the platform covers all graphic data and image data. Users only need to drag and drop to configure all devices that support mobile devices, and make professional analysis reports, dashboards, large-screen data, and graphical interfaces. They can also analyze the complete web page data display system at the same time. The UE big data visualization platform is designed to add unique web page interactive controls to a complete set of web page query systems with differentiated interactive capabilities, which can be used in the design of data display, data acquisition, data linking, data analysis, etc.

(3) Multidimensional data analysis

The power grid government affairs monitoring visualization platform can provide various commonly used OLAP analysis functions. In addition to the most basic data classification and aggregation, orthogonal protection analysis functions can also be introduced. The built-in orthogonal protection data analysis workbench can select any dimension and use any data model as a measurement to realize data analysis of independent transmission and

storage. In the same model, all data can be interconnected, thus providing the freedom of multi-dimensional and timely data analysis. Multi-layer drilling supports the free establishment of hierarchical associations between dimensions and the establishment of multi-dimensional analysis in multi-dimensional analysis. By applying hierarchical dimensions to different charts, automatic multi-level drilling and layer-by-layer analysis can be performed, and different values can be expressed. Concentration is a statistical analysis task that can directly analyze the dimensions and measures of conventional statistics, and examine the overall effects in statistics, including average, maximum, minimum, and general needs. Custom calculation degree, the measurement value can be directly exported in the original data field, but it can also be calculated by formula. The visualization platform supports the creation of custom operation fields. These recently added fields can use one or more original data fields and use common mathematical functions to calculate.

3 Experiment

3.1 Login Module

User authentication function of this module. Compare the data recorded in the database based on the user name and password. If it passes, it will enter the user information module. If it fails, the corresponding reason will be prompted. When the entered user name is invalid or the password is wrong, it will prompt “Invalid user name or incorrect password”. Click Change Password to enter the corresponding password change page. If the original password is entered incorrectly or the new password entered twice is different, it will prompt “The password entered twice is different or the same as the original password”. After the user double-clicks the power grid government affairs monitoring visualization system, he first enters the system login interface. At this time, he must enter his name and password. If you pass the certification, you can proceed to the next step.

3.2 Data Conversion Aggregation Module

Under normal circumstances, companies need to consider operations and analyze data when making decisions. Operating data is dynamic data, which is constantly updated and changed during operation. Analysis data is historical data, static data, and usually does not change over time. The system must be able to migrate from a dynamic data source to a static data source, that is, a typical database to data warehouse conversion is completed. Conversion and data collection can successfully complete this process. If the data warehouse information is inaccurate, the data warehouse will not work. Based on this, only well-designed data can enter the data warehouse. In other words, in order to convert general data into a data warehouse, a data conversion collection system is needed. Generally, the data transformation collection subsystem cleans and filters the data associated with the subject data in the data source before loading it into the data warehouse.

3.3 Comprehensive Query Module

Power grid government affairs involve different tasks and require more data sources. The query of power grid government affairs in this article is based on multi-dimensional conceptual data views. Users can realize multi-angle, multi-level, and multi-dimensional data in a power marketing database. Is not a general query. By modeling and calculating members of different levels and dimensions, users can better understand the essential importance of data. As far as power marketing is concerned, this fast and diversified query method is undoubtedly of great significance.

4 Discussion

System performance monitoring is mainly to quantitatively analyze the performance at the non-functional level to ensure that it meets the expected performance requirements. Professional software system testing tools are usually used for testing and analysis. The LoadRunner automated test tool is mainly used to test the performance of the power grid government monitoring visualization system. The specific test content includes the number of concurrent users of the system, reasonable uptime, server-side server resource occupation, etc. On the test host, multiple users are simulated by running test scripts. Environment, LoadRunner tool reads relevant performance indicators, and after statistical analysis, checks whether the system performance meets expectations.

Table 1. Maximum response time under different concurrent users

	40	80	120	160	200
Login module	1.61	1.75	1.88	1.97	2.03
Data conversion aggregation module	1.77	1.81	1.93	2.00	2.12
Comprehensive query module	1.65	1.76	1.81	1.90	2.04

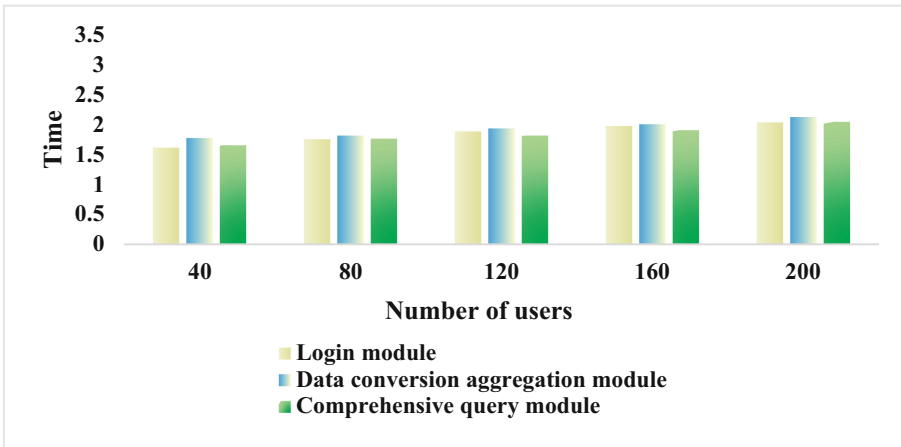


Fig. 1. Maximum response time under different concurrent users

It can be seen from Table 1 and Fig. 1 that under the condition of 5 rounds of pressurization, the maximum response time of the 3 performance monitoring points selected in the performance test work did not exceed 3 s, and the maximum response time was under the condition of the maximum of 200 concurrent users. Under the system, the logical operation of the system still responds normally and meets the expected response time requirements.

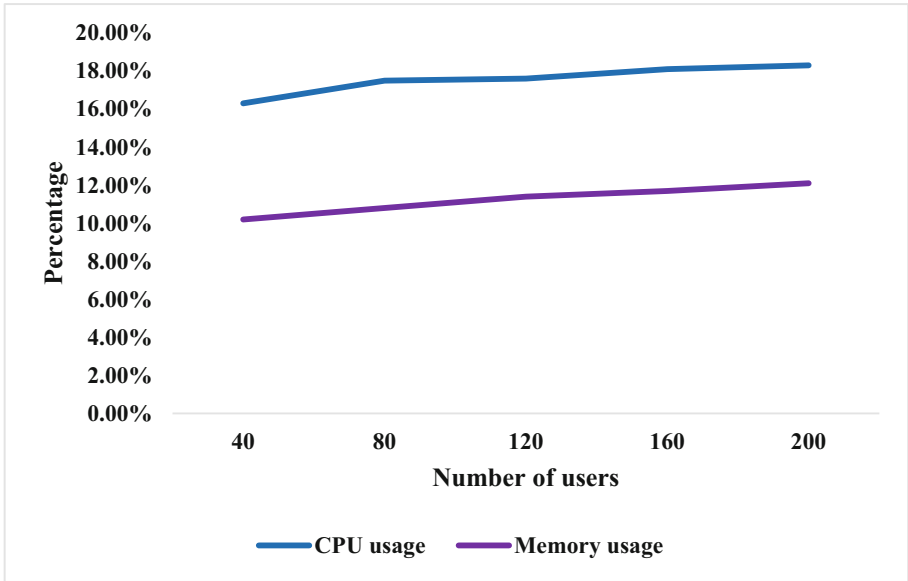


Fig. 2. Server host resource occupancy rate under different concurrent users

It can be seen from Fig. 2 that during the performance test of the power grid government monitoring visualization system, under different concurrent user access pressures, the system's CPU occupancy rate and memory resource occupancy rate did not exceed 20%, and the system was operating stably. Meet the predetermined system resource occupancy rate requirements. Therefore, in summary, the performance test of the power grid government monitoring visualization system can be obtained.

5 Conclusions

The development of the electric power industry is directly related to the smooth progress of people's lives and work, and is also related to the effective development of China's social economy. The power grid government monitoring visualization system has had a relatively profound impact on the development of China's economic and social industries. In the future economic and social development of China, the status of the electric power industry is still critical. It is necessary to improve the national power grid government monitoring and visualization system to enable the healthy development of the electric power industry and drive the overall development of China's social economy.

Acknowledgements. Key Informatization Projects of China Southern Power Grid Co., Ltd (031900HK42200008).

References

1. Qi, J., et al.: The application of improved differential evolution algorithm in electromagnetic fracture monitoring. *Adv. Geo-Energy Res.* **4**(3), 233–246 (2020)
2. Shen, B., et al.: Application of spindle power signals in tool condition monitoring based on HHT algorithm. *Int. J. Adv. Manuf. Technol.* **106**(3–4), 1385–1395 (2019). <https://doi.org/10.1007/s00170-019-04684-0>
3. Li, H.L., et al.: Review on dynamic time warping in time series data mining. *Control Decis.* **33**(8), 1345–1353 (2018)
4. Lee, J., Ohba, N., Asahi, R.: Discovery of zirconium dioxides for the design of better oxygen-ion conductors using efficient algorithms beyond data mining. *RSC Adv.* **8**(45), 25534–25545 (2018)
5. Marozzo, F., Talia, D., Trunfio, P.: A Workflow management system for scalable data mining on clouds. *IEEE Trans. Serv. Comput.* **11**(3), 480–492 (2018)
6. Shi, Y., et al.: Data mining to online galvanic current of zinc/copper Internet atmospheric corrosion monitor. *Corros. Sci.* **133**, 443–450 (2018). <https://doi.org/10.1016/j.corsci.2018.02.005>
7. Zhang, H., Rao, H., Feng, J.: Product innovation based on online review data mining: a case study of Huawei phones. *Electron. Commer. Res.* **18**(1), 3–22 (2017). <https://doi.org/10.1007/s10660-017-9279-2>
8. Denis, G., et al.: The migrate project: the challenges of operating a transmission grid with only inverter-based generation. A grid-forming control improvement with transient current-limiting control. *IET Renew. Power Gener.* **12**(5), 523–529 (2018). <https://doi.org/10.1049/iet-rpg.2017.0369>
9. Klettke, A., Sieberichs, M., Moser, A.: Operational concepts of HVDC in the context of security assessment in the German transmission grid. *Global Energy Interconnection* **2**(02), 40–42 (2019)
10. Klettke, A., et al.: Operational concepts of HVDC in the context of security assessment in the German transmission grid. *Global Energy Interconnection* **2**(2), 130–132 (2019). <https://doi.org/10.1016/j.gloi.2019.07.005>
11. Rivera, J., et al.: Automatic generation of real power transmission grid models from crowdsourced data. *IEEE Trans. Smart Grid* **10**(5), 5436–5448 (2019)
12. Rehhausen, A., et al.: Quality of federal level strategic environmental assessment – a case study analysis for transport, transmission grid and maritime spatial planning in Germany. *Environ. Impact Assess. Rev.* **73**, 41–59 (2018). <https://doi.org/10.1016/j.eiar.2018.07.002>



Cyber Security Practice in Organizations Based on Internet Environment

Jing Lin^(✉)

School of Management, Wuhan Donghu University, Wuhan 430000, Hubei, China

Abstract. The ever-changing Internet technology had not only improved the breadth and speed of interpersonal communication, but also provided people with a broad and efficient stage for activities, which brought convenience to us. But on the other hand, it also had created multiple pressures and challenges for cyber security governance. Three types of enterprises were selected to conduct a survey on cyber security awareness and technology level, and a comprehensive model of cyber security practice integrating ideology, talent training, legal restriction, process monitoring and technological innovation was proposed on the basis of the investigation results.

Keywords: Cyber security practice · Network awareness · Internet environment · Comprehensive model

Since the 18th National Congress of the Communist Party of China, a series of new judgments, viewpoints and requirements had been put forward for network ideological governance on the base of accurately grasping the new characteristics and rules of the development of network society and paying high attention to the profound impact of cyberspace on all fields of the world. They contained not only the cognition of pressure and challenges which was brought by the ideology of multiple network to the national governance, but also the goal orientation, fundamental requirements, basic principles and specific strategies. It provided an important basis for establishing the network ideological governance system and realizing the modernization of governance capacity under the new situation [1].

1 Problems of Cyber Security in Organizations

1.1 Internal Aspects of Organization

1.1.1 Lack of Safety Awareness

The safety consciousness of all employees was the first firewall that constituted the cyber security of organization, even if a small gap might lead to extremely serious consequences. The reality, however, was that it was difficult to fully understand the importance of cyber security in organizations from managers down to low-level employees. They might use their work computers to browse webpage at will or upload internal data of organization to the cloud without security verification. After conducting a questionnaire

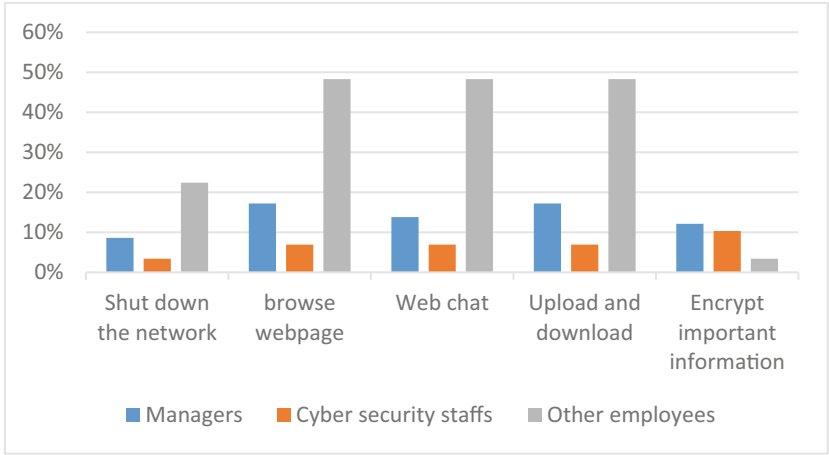


Fig. 1. Results on employees' awareness of cyber security

survey on cyber security among employees of some organizations, this paper drew the following conclusions in terms of employees' security awareness.

The survey covered three enterprises, including a bank, a subsidiary of a state-owned enterprise and a private enterprise. Sixty people were investigated, and 58 effective samples were recovered with an effective recovery rate of 96.7%. Among the 58, 15 were managers and 43 were junior staffs, including six who were responsible for cyber security. As can be seen from Fig. 1, none of the bank employees' office computers were connected to external network for the sake of security. In the other two companies surveyed, however, employees were not restricted from using office computers to surf the Web, use chat tools, upload and download files or other online activities. Only cybersecurity employees of the surveyed organizations and managers of the bank had paid more attention to encrypting important information.

1.1.2 The Level of Security Application in Organizations Was Uneven

The professional skills of Internet technical personnel engaged in development, testing, operation and maintenance in organizations were qualified, but the improvement of the security level of other employees was often ignored. The frequency of cyber security training varied greatly among the three surveyed organizations, and the participants were unevenly distributed. The bank carried out cybersecurity training every quarter for all its employees, in addition, its cyber security staffs were also arranged to attend cyber security training courses outside. The state-owned enterprise arranged cyber security training for all employees once a year, and the private enterprise only arranged cyber security staffs to participate in cyber security training outside timelessly. In the face of network threats, many employees were unable to identify or control them effectively, which would lead to serious consequences ultimately.

As shown in Fig. 2, professional cyber security personnel had a better grasp of professional skills than managers and other employees. But due to the lack of actual

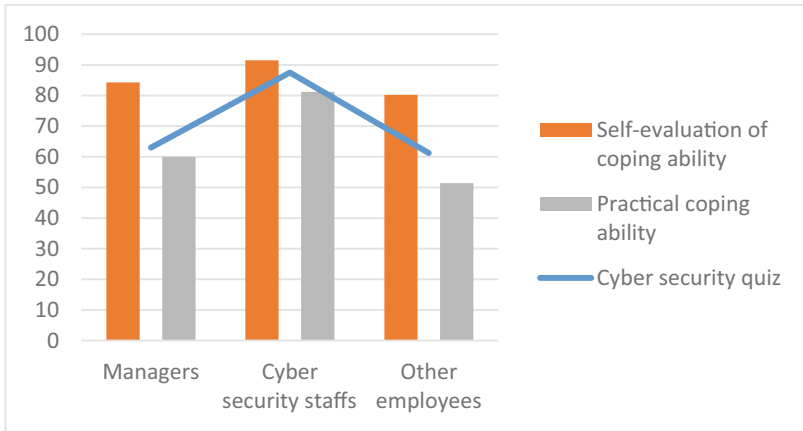


Fig. 2. Results on employees' cyber security capabilities

combat exercises, all of them had overestimated their level, while managers and non-cyber security employees were more obvious.

1.1.3 The Professional Level of Cyber Security Needs to Be Improved

Most organizations lacked professional core security personnel, which made the organization very passive when it faced security threat, and this situation was particularly significant in private enterprises. In this investigation, it was found that the cyber security employees in bank had better professional ability, and they were proficient in API attack detection, network anomaly detection, network security situation awareness and

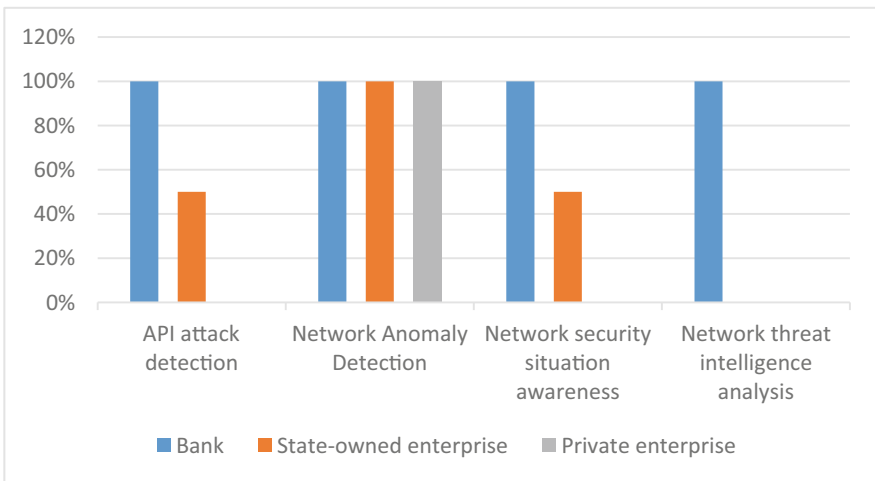


Fig. 3. Results on the professional level of cyber security

network threat intelligence analysis, which were obviously superior to the state-owned enterprise and private enterprise.

1.2 External Aspects of the Organization

1.2.1 Internet Rumors

Because of the characteristics of fast propagation, destructive power and difficult investigation, Internet rumors had brought challenges to organizational cyber security. Before the official released the information timely, Internet rumors become an important basis for people to obtain and judge information. It would not only cause serious harm to innocent people, but also pollute cyberspace if unchecked.

1.2.2 Collision of Ideas

With the popularization of the Internet, all kinds of ideas spread and collided on it, in particular, some hostile forces used the Internet to spread bad ideas, which had seriously broken the ideological security of the Internet. If the organization lacked advanced network threat monitoring technology and means, especially the application of all kinds of big data technology, it would cause the invasion of bad ideas and bring bad consequences.

1.2.3 Defects in the Platform

Network platform not only expanded traditional forms of information transmission, but also lead to frequent social adverse events with its own characteristics of openness and virtuality. Network crimes were low in cost, high in concealment and difficult to trace, these loopholes caused by the defects of the platform were full of network and threaten the cyber security.

1.2.4 Technical Constraints

Network technology changed with every passing day, the maturity of technology had brought about changes in this field. However, the absolute discourse power and technological advantages in this field had been always held by western countries. They could control the production of online content by virtue of their technological advantages, pushed targeted information to users in various ways, and manipulated the power of online discourse. In this respect, we had been frequently hampered.

2 New Opportunities to Govern Cyber Security in Organizations

The security of organizational network ideology was directly related to that of national ideology, and the safeguard of organizational network ideology security was an important means to maintain national security, which was related to the right of discourse, guidance and supervision of China's ideological work [2]. At the same time, it was also an inevitable requirement for the overall development of employees. The safe and

healthy enterprise network environment would help employees establish correct ideology. These “chilled culture”, “lie down” and other non-mainstream network ideology seriously damaged the good growth environment of employees.

With the emergence and development of network information exchange platform, the efficiency of cyber security governance had been also greatly improved. Organizations can use the Internet to carry out relevant subject education for employees, so as to cultivate their awareness of cyber security. In addition, the two-way communication on the network was also conducive to stimulate the enthusiasm of staffs to participate in cyber security management.

3 Approach for Cyber Security Practice in Organizations

3.1 Building the Systematic Internet Governance System

The key of cyber security was to establish a comprehensive early-warning system on organizational cyber space. Among which, the early stage was security warning mechanism, the middle was network threat analysis and guidance mechanism, and the late was network threat sharing mechanism. At the same time, relevant network laws and regulations should be improved to limit and standardize the network access permissions and rules of different employees. A special leading group of network public opinion management should also be set up to respond to enterprise network emergencies and improve the level of cyber security prevention. In addition, strengthening international exchanges and cooperation had been also the general trend of Internet construction. Cyber security is a common challenge faced by all human society. Therefore, cooperation and information sharing with other countries should be strengthened to jointly combat cybercrimes.

3.2 Constructing the System of Network Talent Team

Cyber security talents are the core resources for the construction of cyber power, and their quantity, quality, structure is directly related to the capability in cyberspace security [3]. But the higher education system lacked the training of Internet governance talents in our country, in contrast, the United States and Japan and other cyber power had already started in this aspect. For example, in order to safeguard its cyberspace strategy, the United States had established a complete training and education system and trained a large number of information technology professionals [4]. While vigorously developing information education, Japan accelerated the training of information security technical personnel to improve the strategic talent structure in cyberspace [5]. Therefore, we must develop innovative ideas and concepts for the construction of cyber space security personnel relying on the whole social forces and strengthen the awareness and training of cyber security to form a multi-level talent system of cyber power.

3.3 Breaking Through Core Network Technologies

It is necessary to have effective tools to do good work. Core technology controlled by others had been our biggest hidden danger. Therefore, it is necessary to increase investment in scientific and technological research, break the dependence on other countries,

and promote the construction of cyber security with technology. First of all, in terms of hardware and software, the investment in scientific research should be increased to form our own technical standards and device production system, we need to build the advantages of localized network information technology [6]. Secondly, we should pay attention to the development of network monitoring and defense technology, establish big data analysis models to improve the capabilities of data collection and analysis and the network security by detecting, filtering, and eliminating malicious attacks in a timely manner.

To sum up, the realization of cyber security in the Internet environment needs to be based on the ideology and the concept of cyber security, rely on cyberspace talent training and cyber security laws, establish a monitoring system with safety precaution, threat analysis and public opinion guidance as the main body, and break through the core network technology of hardware and software and defense technology at the same time. All of them work together to advance cyber security practices with the sharing and cooperation of international cyber threat and development information.

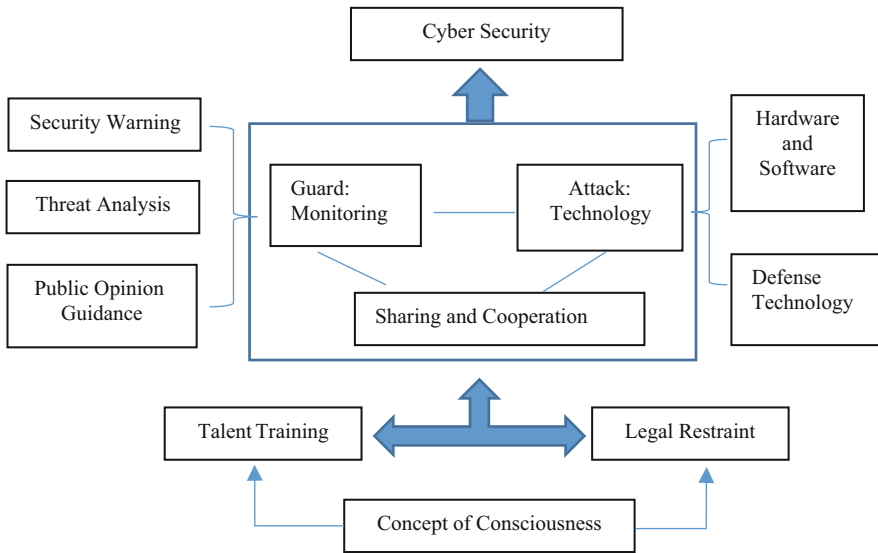


Fig. 4. The mode of cyber security practice

4 Conclusions

At present, the cyber security situation is still not optimistic. In terms of technology, we still lacked advanced methods of network threat and attack detection at the present stage. At the same time, the shortage of cyberspace management talent brought us big hindrance [7–10]. In the follow-up study on cyber security and intelligence analysis, it is necessary to further intensify the research on big data technology under the condition

of cyberspace security, seek the containment methods of network attacks under complex conditions, and do a good job in risk perception, detection, early warning, intelligence sharing and other aspects.

Acknowledgements. This work was supported by the grants from Youth Foundation Project of Wuhan Donghu University (project number: 2020dhsk003).

References

1. Wang, Y., Yue, A.: Focus on building a clear and clean cyberspace – study on the important discussion of chairman Xi’s network ideology governance thought. *J. South Cent. Univ. Nat. (Humanities and Social Sciences)* **37**(04), 1–8 (2017)
2. Sozansky, T., Krasnytskyi, I.: International practice of legal support of cyber security of the country. *J. Leg. Ethical Regul. Issues* **23**(2), 1–8 (2020)
3. Vaseashta, A., Susmann, P.: Strengthening cyber security policy by means of end-users dedicated training. *NATO Sci. Peace Sec. Series-D Inf. Comm. Sec.* **6**(38), 49–56 (2014)
4. Sarker, I.H., Kayes, A.S.M., Badsha, S., Alqahtani, H.: Cybersecurity data science: an overview from machine learning perspective. *J. Big Data* **7**(01), 243–259 (2020)
5. Gibert, D., Mateu, C., Planes, J.: The rise of machine learning for detection and classification of malware: research developments, trends and challenges. *J. Netw. Comput. Appl.* **153**(C), 102526 (2020)
6. Vardi, M.Y.: Ten years at the helm of communications of the ACM. *Commun. ACM* **60**(06), 34–38 (2019)
7. Janet, C.: Improve cyber security and protect practice finances. *Med. Econ.* **92**(24), 29–33 (2015)
8. Oyelami, J.O., Kassim, A.M.: Cyber security defence policies: a proposed guidelines for organisations cyber security practices. *Int. J. Adv. Comput. Sci. Appl.* **11**(08), 223–226 (2020)
9. Simone, H.-L.: Cyber security: protecting client confidentiality in the digital era. *LSJ: Law Soc. NSW J.* **6**(56), 88–89 (2019)
10. Simone, H.-L., Gabor, S.: Cyber security: owning your practice’s cyber security: MSPs and you. *LSJ: Law Soc. NSW J.* **2**(52), 80–81 (2019)



Analysis of the Impact of New Energy Access on Smart Distribution Network Operation Based on Deep Learning Technology

Guihua Qiu^(✉), Zhirui Tang, and Zijia Kuang

Foshan Power Supply Bureau of Guangdong Power Grid Co., Ltd., Foshan 528000, China

Abstract. In order to objectively understand the effects of new energy access for smart distribution network operation, this article, taking the analysis of deep learning technology, this paper mainly introduces the basic concept of deep learning technology, new energy and the relationship between the intelligent distribution network, the combined with deep learning technology to establish relevant system, and then through the system to analyze the influence of the new energy access, finally puts forward relevant Suggestions. The system in this paper can be used to analyze the influence of new energy access, and then the influence can be eliminated to a large extent through strategies, so that new energy can be smoothly connected to the smart distribution network, which is more conducive to the role of new energy.

Keywords: Deep learning technology · New energy sources · Intelligent distribution network

1 Introduction

At present, China began to apply intelligent technology to the operation of the distribution network, improving the management level of the operation of the distribution network. Because there are some negative effects on the use of the energy required by the traditional electric power production, many new energy sources that can be used for electric power production have been introduced after years of development. This case, the country began to carry out the combination of new energy and intelligent power distribution network, and with the help of powerful management functions of intelligent terminal in intelligent power distribution network for management, can fully improve the quality of distribution network operation efficiency, in theory, but the reality is a new energy access after the intelligent distribution network or caused some impact, in the face of these effects must be taken corresponding management methods, Therefore, it is necessary to carry out relevant research and analysis, aiming at clarifying the form of impact, and then combining with intelligent terminals to set up management methods to eliminate the impact.

2 Basic Concepts of Deep Learning Technology and the Relationship between New Energy and Smart Distribution Network

2.1 Basic Concepts

Deep learning technology is a systematic technology, including big data, intelligent technology and neural network technology. Its basic operation mode is as follows: First of all, on the basis of the large data information of the original data, construct the corresponding neural network model, secondly by intelligence technology makes neural network model identification of information data, computing power, the neural network model will according to the data identification and calculation results, learning continuously, the process of deep mining, the original information of huge data So this way is known as the “deep learning”, synchronous every learning process and results are recorded to the knowledge base and finally when encounter information technology systems in the operation data, in the knowledge of the rolls to find if there is a corresponding record, if you can analyze the information data represented by the event, the objective and reasonable multiple decision-making advice, Manual can make corresponding choices according to actual needs, otherwise if there is no record, then a new round of deep learning will be carried out. From here you can see, the depth of the learning technology has a high application value, to be able to help people analyze all kinds of problems, and will series every information in the process of analyzing data, under the background of the current information data in the human is unable to do this, even if force analysis, the result is often incomplete, such as partial subjective quality problems, However, the results obtained by deep learning technology are more complete, objective and reasonable. In addition to this, deep learning technology also has a strong domestic sex, namely the main function of the technology is the information data processing, analysis, operation itself is not restricted by information data types, sources, at the same time from the modern perspective, the reality of everything is the source of information data, so deep learning techniques can be applied in many fields, Including electric power transmission and distribution field. It is worth mentioning that neural network is the basis of the three technologies in deep learning technology system, and its basic model is shown in Fig. 1.

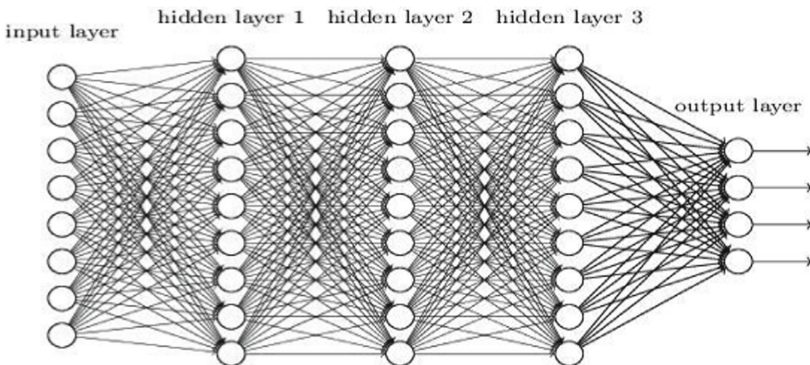


Fig. 1. Basic model of artificial neural network

Can be seen in the figure, the artificial neural network can be divided into input layer, hidden layer and output layer, the input layer of each dot represents a primitive information data production of neuron node, each neuron node and link between internal neurons of hidden layer nodes, shows that artificial neural network will be starting from each original neuron node, Numerous attempts are made in hidden layers. There is no limit to the number of hidden layers, depending on the relationship between the data. At the same time, the neuron nodes in each hidden layer are also not limited and will expand indefinitely according to the needs of information data analysis until data analysis enters the end point and corresponding analysis results are obtained, which will be displayed in the output layer. In addition, there are many data analysis methods about artificial neural network, among which the more representative method is particle swarm algorithm, see formula (1), (2), (3) for details [1–3].

$$x_i = (x_{i1}, x_{i2} \cdots, x_{iD}) \quad (1)$$

$$u_i = (u_{i1}, u_{i2} \cdots, u_{iD}) \quad (2)$$

In the formula, x represents the position of the particle, u represents the speed of the particle, and I is the value of the position or speed, with a value range of [1, 2]. The algorithm parameters can be initialized through the two formulas, making preparations for formula (3).

$$u_{ij} = \omega u_{ij} + c_1 r_1 (p_{ij} - x_{ij}) + c_2 r_2 (p_{gj} - x_{ij}) \quad (3)$$

Where: $j = 1, 2, \dots, D$; c_1 and c_2 are learning factors. r_1 and r_2 are random numbers between [0,1]. P_{ij} is the individual optimal position of particle I , and P_{gj} is the global optimal position. The global optimal solution can be obtained through formula (3), which represents the end of a single neural network learning. After that, it will continue to cycle in the same way until all neuron nodes of original information data have corresponding optimal solution.

2.2 Relationship

New energy is a new type of energy that is different from traditional electric energy production. It generally refers to light energy, wind energy, water energy, etc. (traditional energy refers to thermal power generation). Compared with traditional energy, these energy have many advantages, but also some disadvantages: First, in terms of advantages, new energy generally has the advantages of high cleanliness, easy access, strong reproducibility, high recycling value, low cost, no pollution, etc., such as light energy has all the advantages of characteristics, so China is vigorously implementing photovoltaic power generation, which has now become China's representative technology in the production of new energy energy; Second, although new energy can be used for electric energy production, the production efficiency is insufficient. Under the same time and consumption condition, the electric energy produced by new energy may be less than half of that of traditional energy. From this point of view, the main source of electric energy in modern intelligent distribution network is still traditional energy. However, faced with the huge demand of electric energy across the country, the efficiency of

electric energy production of traditional energy is also insufficient. If we continue to use traditional energy alone for electric energy production and then carry out transmission and distribution, the overall demand of electric energy cannot be met. Constantly at the same time increase the input of traditional energy can cause large economic pressure to the state, also the severity of the problem such as environmental pollution, in the home so need based on the traditional energy utilization of new energy to fill the gaps, so the new energy in the present mainly as a secondary energy access smart electricity production and distribution of the distribution network link, it is the embodiment of the relationship [4, 5].

3 Deep Learning Technology Design and Impact Analysis

3.1 System Design

Figure 2 is the basic framework of the deep learning technology system in this paper.

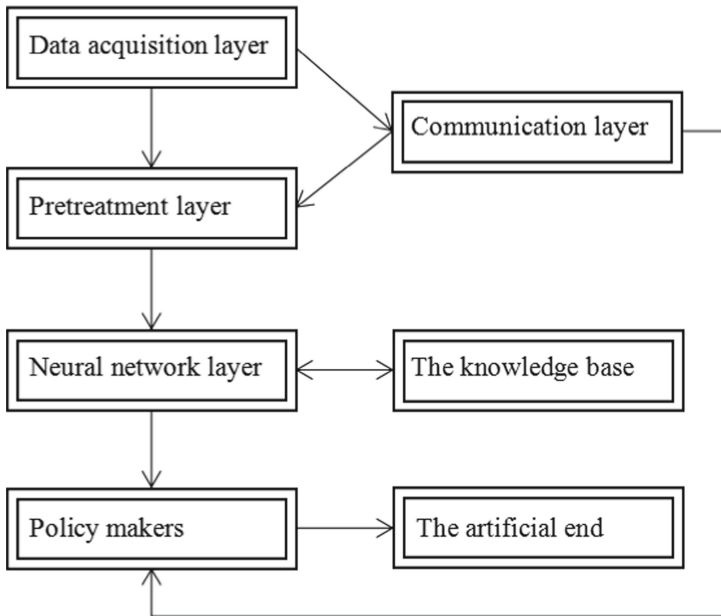


Fig. 2. Basic framework of deep learning technology system

The system first obtains the information data of the distribution network affected by new energy access through data acquisition equipment. The data information will be converted into digital format through the transducer, and then sent to the communication layer, through the communication layer into the pretreatment layer. Second in the preprocessing layer, with the aid of preprocessing tool to get the original information of data standardization comparison, remove the incomplete, don't conform to the requirements specification information data, to ensure the purity of information data,

preprocessing available after the completion of related data, the data will be postpone system into the artificial neural network. Again, the artificial neural network will cycle deep learning according to the intelligent logic in the knowledge base (the intelligent logic needs manual programming input, which will not be described here). The learning process and results will be returned to the knowledge base, so that the results can be directly given next time, and the synchronous learning results will be sent to the decision layer. Finally, the decision layer also keeps contact with the communication layer, and under the support of the communication layer, all the decision suggestions will be transmitted to the artificial end, and the artificial end can understand the influence of new energy access by making a choice.

3.2 Impact Analysis

With the help of deep learning technology system, this paper analyzes the impact of new energy access on smart distribution network, and understands that new energy access will cause three impacts on smart distribution network, as follows.

3.2.1 Power Quality is Affected

Purely relying on traditional electricity intelligent distribution network although cannot fully meet the requirements of national power, but its power quality can be guaranteed, but after new energy access, as part of the new energy easily affected by external factors, thus causes the electric energy production often fluctuate, short-term stagnation phenomenon, such as wind power, photovoltaic is this kind of phenomenon, At the same time, the new energy power generation is equipped with rectifier inverter equipment and other power electronic equipment, but also produces harmonic and DC components. In this case, the electric energy generated by the new energy is connected to the smart distribution network, which is easy to cause the voltage fluctuation of the power grid, which makes the power quality unstable, or affects the measuring instrument, which makes it unable to measure accurately, or leads to the misoperation of relay protection. In addition, because the new energy generally is through the access of the intelligent distribution network interconnection way, so the fluctuation characteristics of electricity production will result in power distribution network and volatility, this affects power electricity regulatory, actual in the instant access to a larger impact current, high voltage fluctuation probability distribution network caused by wire, the phenomenon such as discharge, The situation is serious and even directly lead to the burning of lines and equipment. In general, the access of new energy to smart distribution network is not smooth, which will adversely affect the power quality in the distribution network [6–9].

3.2.2 Increase Line Loss

Line loss has always been an unsolvable problem in power transmission and distribution, but with the long-term development of technology, the problem has been reasonably controlled and the degree is acceptable. However, the access of new energy makes the line loss problem out of control to a certain extent, so new energy causes the impact of increasing line loss. In principle, when new energy access intelligent distribution

network, distribution system circuit network will be from the original single power radiation form to users connected electroweak ring form, this change will lead to the power distribution in the form of the distribution of load size, distribution and transmission direction become unpredictable, so you cannot effectively control, lead to increase the line loss problem [10]. Figure 3 shows the mechanism of increasing the influence of line loss on new energy access.

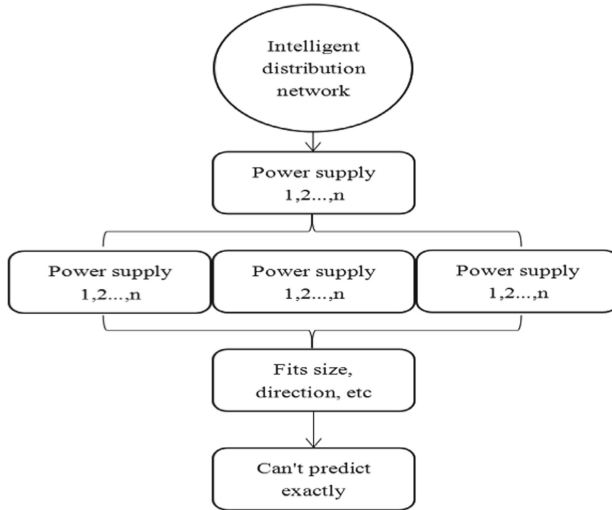


Fig. 3. Mechanism of new energy access increasing the influence of line loss

3.2.3 Adverse to Real-Time Monitoring

Intelligent distribution network operation is a dynamic process, in the process of electricity all the time change, so to do a good job in the corresponding management must be real-time monitoring of the process, through the sensor devices such as basic requirements, press 1 min/time cycle data acquisition, transmission, according to the data grid, find exceptions can be handled in time. But after new energy access, ever real-time monitoring system have been affected, the new energy access process is more complex, part of the power source may be ignored, it'll be easier to form power island phenomenon, but electricity and really into the intelligent distribution network, thus became an island of electricity is not monitoring, through real-time monitoring work, it is difficult to accurately people The results produced are often inaccurate. Island at the same time, the electricity is not monitored, so it caused some abnormal problems easy to affect the user experience of electricity, and even threaten the safe use of electricity, such as the islands of the voltage of electricity may overlocking, client some of the appliances easily burned at this moment, if the user is using the device, it is likely to cause safety accidents.

4 Suggestions for Dealing with Impacts

Through analysis on new energy access to a variety of effects on the intelligent power distribution network, the currently no effective treatment, may also affect early but theoretically have the related countermeasures and Suggestions, such as the electric power enterprise can plan period of traditional and new energy into the electricity grid, which belongs to the peak season during the day, so in this phase can give priority to with traditional electricity power supply, In this way, the new energy is mainly intermittent power supply during the day, while the electricity consumption is in the trough at night. At this time, the electric energy accumulated by the new energy equipment during the day can be used as the main output, while the traditional electric energy can be temporarily stopped. In order to achieve this, it is suggested that electric power enterprises establish electric power supply line switching system, with the help of this device can reduce the influence of new energy to a certain extent. Figure 4 shows the principle of power supply line switching system.

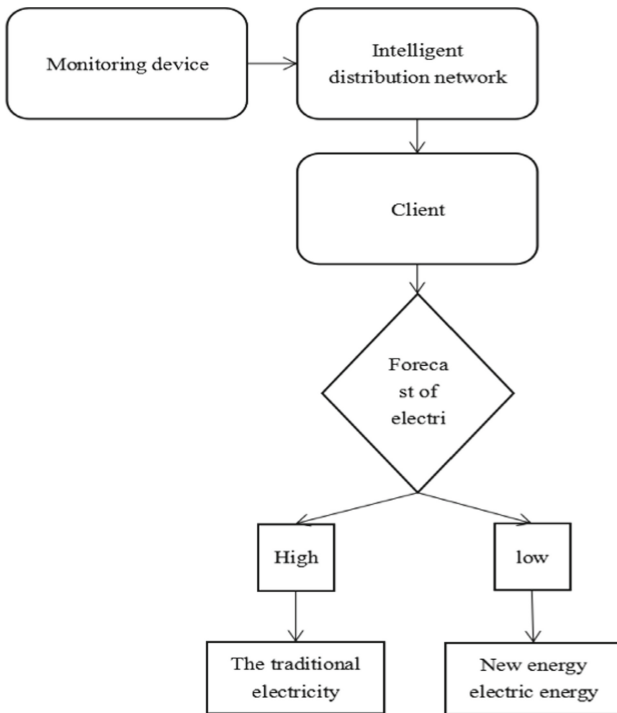


Fig. 4. Electric power supply line switching system principle

5 Conclusion

To sum up, new energy access is imperative, but its impact on smart distribution network should not be ignored. Therefore, deep learning technology is needed to understand the

impact of new energy access, so as to find ways to eliminate the impact through subsequent technology research and development. At the same time, because the influence of the current access of new energy cannot be completely eliminated, it can be controlled and avoided as much as possible with the help of some theoretical suggestions, which can at least further play the role of new energy.

References

1. Ustun, T.S., Aoto, Y.: Analysis of smart inverter's impact on the distribution network operation. *IEEE Access*. 1 (2019)
2. Kumar, D.S., Savier, J.S.: Impact analysis of distributed generation integration on distribution network considering smart grid scenario. *IEEE*, 1–5 (2017)
3. Possemato, F., et al.: On the impact of topological properties of smart grids in power losses optimization problems. *Int. J. Electr. Power Energy Syst.* **78**, 755–764 (2016)
4. Bigot, J., et al.: An approach to increase reliability of hpc simulation, application to the GYSELA5D Code. *Esaim Proc. Surveys* **53**, 248–270 (2016)
5. Dempwolf, C.S., Lyles, L.W.: the uses of social network analysis in planning: a review of the literature. *Journal of Planning Literature Incorporating the CPL Bibliographies* **27**(1), 3–21 (2012)
6. Wattanakuljarus, A., Coxhead, I.: Is tourism-based development good for the poor? *J. Policy Model.* **30**(6), 929–955 (2008)
7. Afify, E.A., Eldin, A.S., Khedr, A.E. Facebook profile credibility detection using machine and deep learning techniques based on user's sentiment response on status message. *Int. J. Adv. Comput. Sci. Appl.* **11**(12), 622–637 (2020)
8. Stuhlsatz, A., Lippel, J., Zielke, T.: Feature Extraction with deep neural networks by a generalized discriminant analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **23**(4), 596–608 (2012)
9. Mead, P.E.: Evolutionarily repurposed networks reveal the well-known antifungal drug thiabendazole to be a novel vascular disrupting agent. *Plos Biol.* **10**(8), e1001379 (2012)
10. Greu, V.: Using the information and communications technology data deluge from a semantic perspective of a dynamic challenge: what to learn and what to ignore?-part 3-. *Romanian Distribution Committee Magazine*. 11 (2020)



Application of “Blockchain + 5G” Technology in the Transformation of Financial Digital Intelligence

Yanqing Chen^(✉)

College of Accounting, Zhanjiang University of Science and Technology,
Zhanjiang 524000, Guangdong, China
20162103779@mails.imnu.edu.cn

Abstract. Blockchain technology is one of the representative technologies of technological innovation in the new era. It has the advantages of distributed ledgers, collective maintenance, reliable databases, non-tamperable, and traceable, and has gradually integrated into various industries in the economic field. Finance is the business language of social economy. The combination of “blockchain + 5G” technology applied in the financial field can promote the transformation of finance to digital intelligence. This article mainly aims at the application research of “blockchain + 5G” technology in the transformation of financial digital intelligence. This article analyzes the application of “blockchain + 5G” technology in the intelligent storage of financial statement data, corporate financial internal supervision and auditing, and capital management. This paper uses a case study to analyze the application effect of the group’s “blockchain + 5G” technology application financial system. Data shows that M Group built and started to use a financial analysis platform based on “blockchain + 5G” technology in March 2017. The order tracking rate increased by 7.46%, the after-sales collection rate increased by 3.34%, and the cost accounting rate increased by 9.68%. It can be seen that the application of “blockchain + 5G” technology in the financial field has a good effect.

Keywords: Blockchain technology · Financial data intelligence · Accounting · 5G technology

1 Introduction

Since the 21st century, the global economic structure and technological ecology have been reconstructed [1, 2]. The new generation of technical theoretical research represented by blockchain, Internet of Things, artificial intelligence, quantum computing, 5G communications, etc. has accelerated and gradually broke through practical applications [3, 4]. The application of “blockchain + 5G” technology in the financial field has promoted the gradual transformation of this field to intelligent data storage, digital asset transactions, and digital authentication [5, 6].

Regarding the research of blockchain technology, many scholars at home and abroad have conducted multi-angle discussions on it. For example, Lee J studied the financing channels of artificial intelligence supervision in the financial service industry [7]; Begeau J used financial big data as research object, analyze the outstanding effects of big data application companies [8]; Wang L conducts research on the value creation of blockchain-driven supply chain in finance [9]. It can be seen that it is particularly important and urgent to further explore and study the application of “blockchain + 5G” in the financial field, so that financial work can keep up with the entire scientific and technological ecological environment.

The research object of this article is the application of “blockchain + 5G” technology in the transformation of financial digital intelligence. The purpose of the research is to focus on the development of new technologies and new information the application advantages and breakthroughs of technology in the financial field, exploring how to make “blockchain + 5G” and the financial field achieve better integration, and help the economic development of the new era.

2 Application of “Blockchain + 5G” Technology in the Transformation of Financial Digital Intelligence

2.1 Intelligent Storage of Financial Statement Data

(1) Distributed encrypted access

In the financial system under the “blockchain + 5G” technology environment, any transaction data will be stored in a database in real time, the database is shared, and usage data can also be found at any time. In addition, this data chain uses encryption algorithms, which can be entered at any time and encrypted at any time, and only authorized personnel within the company can enter, and the external financial information needs cannot be obtained, which strengthens the confidentiality of the data.

Similarly, companies and any customers and sales ends can correct and analyze financial information such as main business costs, accounts receivable, bad debt preparations, etc. at any time, so that financial personnel can provide financial analysis to management personnel in a timely manner to help the company do Make strategic decisions.

(2) The credible traceability system of financial information

The blockchain adopts a distributed network structure, and the ledger data is stored in each node of the entire network. Tampering or attacking a single node will not affect the normal operation of other nodes, and only when more than half of the nodes are attacked will it cause data loss or modification [10, 11]. Blockchain also has the characteristics of openness and transparency. Except for specially encrypted private information, all system information is highly transparent. Each node can check and verify the authenticity and reliability of global data, which improves the traceability of the financial system.

Specifically, using this feature of the blockchain can connect personnel with financial information traceability requirements in a system, store financial information in the blockchain, and relevant personnel can be used to initiate traceability

of financial information through the traceability source code. Function to solve the problem of credible traceability of corporate financial information.

2.2 Internal Supervision and Audit of Corporate Finance

(1) Simplify the supervision and audit process

The distributed ledger under the “blockchain + 5G” technology has its own monitoring function to ensure the integrity, reliability and security of enterprise assets at any time. In terms of internal control, for each financial business transaction, when recording the complete transaction data of each block, all original vouchers related to the business are registered and stored in an electronic version [12].

(2) Reduce supervision and audit costs

In the Internet age, the Internet audit model is becoming more and more popular. However, Internet audit has a number of costs, such as management and control software maintenance fees, network resource search fees, and manual management costs. With the continuous development of the Internet, the cost of network auditing continues to increase. However, under the “blockchain + 5G” technology, the preparation time for the audit will be greatly shortened. When the company’s economic activities occur, the evidence required for the audit is already stored in the blockchain database. Internal supervisors and external reviewers only need to call out the data to make an analysis. There is no need for spot checks and large amounts of verification. Therefore, while saving labor costs, the “blockchain + 5G” technology also increases the intensity of audit verification. Through real-time review and supervision, time costs are reduced and work efficiency is improved.

(3) Improve the independence of internal audit

The hash encryption algorithm and the AES symmetric encryption algorithm are applied in the blockchain, which can be used to prove the integrity and accuracy of the data. The mathematical expression of the hash function is shown in formula (1):

$$h = H(m) \quad (1)$$

Among them, h represents a fixed-length hash value, H represents a hash function, and m represents an input of any length. The AES symmetric encryption algorithm belongs to block encryption. The key and data block can be selected in lengths such as 128 bits and 256 bits. The principle is based on iteration, replacement and substitution. Specify the AES key derivation function PBKDF (Password-Based Key Derivation Function, password-based key derivation function), as shown in formula (2):

$$DK = PBKDF(P, S, i, len) \quad (2)$$

Among them, DK represents the derived AES key; $PBKDF$ represents the key derivation function, such as Scrypt, P represents the password, S represents the salt value, i represents the number of iterations, and len represents the length of the DK .

The combination of the two algorithms allows any block to have its own unique hash value. Once a piece of data is tampered with, the root hash generated at the end of this block is absolutely impossible to be consistent with the root hash of other nodes. Therefore, the synchronization cannot be completed. This public key algorithm in the blockchain is added to the technical layer to enhance the security of financial data. At the same time, the application of cryptographic algorithms in blockchain technology has formed a supervision and containment function within the company, thereby improving the quality of internal control and supervision.

2.3 Advantages of Fund Management Under “Blockchain + 5G” Technology

(1) Advantages of investment fund management

In the world of “blockchain + 5G”, all market transactions and transactions between enterprises have the characteristics of coding and proceduralization. Code is the law. Enterprises and government organizations are no longer pyramids, but distributed organizational structures, and the power is not decentralized and distributed on the nodes of each blockchain. Organizational operation is also a distributed and highly autonomous market entity, which ensures the authenticity and reliability of market information and eliminates the problem of information asymmetry. And because all information can be interconnected under the blockchain conditions, financial institutions can have comprehensive and complete corporate information, and investors can borrow at a fixed interest rate without risk.

(2) Corporate operating fund management

Enterprises collect information, analyze data, and discover the causes of problems through internal control. The “blockchain + 5G” technology provides the largest data resources in terms of information authenticity and completeness. Moreover, there is no need to review the authenticity of the data again, and the time record of each business provides an activated distributed ledger for the capital block ledger. The fund account information is updated in a timely manner, cannot be tampered with, and is complete and reliable. This allows the financial management of funds to be intelligently reviewed, and illegal use of funds will not be recorded in the blockchain, and such transactions will also be rejected.

3 Research on Experimental Design of “Blockchain + 5G” Technology in the Transformation of Financial Digital Intelligence

3.1 Research Objects

Taking the financial analysis system of M Group as an example, the group has explored the application effect of the traditional financial analysis model and the construction of the financial analysis system of “blockchain + 5G” technology.

3.2 Data Sources

The data of this research comes from the company's annual public data, and the data in this article is formed after collation.

3.3 Research Process

In March 2017, M Group built a financial data mining and analysis platform based on "blockchain + 5G" technology on the basis of the original data collection system and accounting information system. Therefore, this article collects relevant data on the work of financial staff from 2015 to 2020, including the number of financial reports received by the group each month and the time limit for the group to complete the analysis report.

4 "Blockchain + 5G" Technology in the Financial Digital Intelligence Transformation of Experimental Design Analysis

4.1 Benefit Analysis

M Group built and started to use a financial analysis platform based on "blockchain + 5G" technology in March 2017, and it has received good results since it went online. As shown in Table 1, the order tracking rate has increased by 7.46%; the after-sales collection rate has increased by 3.34%; the cost accounting rate has increased by 9.68%; the budget estimation rate has increased by 4.88%; and the material utilization rate has increased by 3.81%.

Table 1. Benefit analysis (%)

Project	Before improvement	After improvement	Improve
Order tracking rate	86.11	93.57	7.46
After-sales collection rate	95.21	98.55	3.34
Cost accounting rate	75.01	84.69	9.68
Estimated budget rate	90.90	95.78	4.88
Material utilization	90.90	94.71	3.81

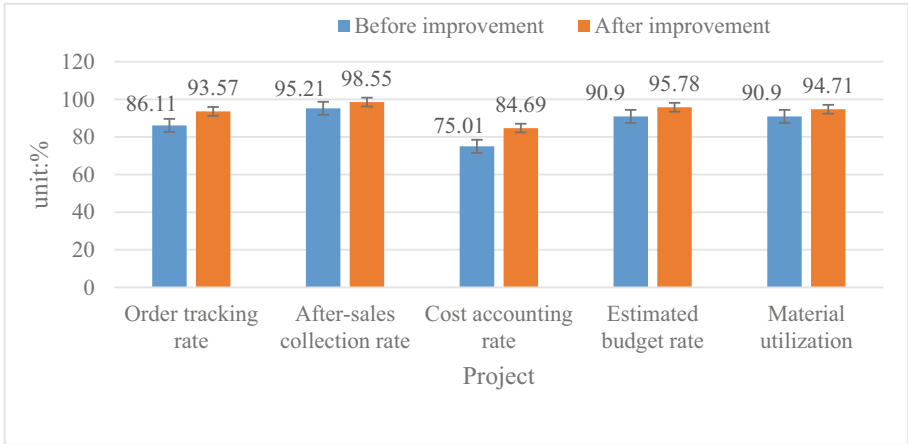


Fig. 1. Benefit analysis (%)

It can be seen from Fig. 1 that the application of related technologies has obviously improved the cost accounting rate and order tracking rate of M Group, saved costs for the enterprise, and improved the management and decision-making efficiency of the group.

4.2 Financial Analysis Efficiency

The work efficiency of financial staff is related to the number of financial reports received and the timeliness required to complete the analysis report. The data is shown in Table 2: In 2016, the number of financial reports received by the group every month was 306, and it took 19 days for the group to complete the analysis report.; In 2017, the number of financial reports received by the group per month was 329, and it took 16 days for the group to complete the analysis report; in 2017, the number of financial reports received by the group each month was 334, and it took 5 days for the group to complete the analysis report.

Table 2. Financial analysis efficiency

Years	The number of financial reports received by the group each month (copy)	Time limit for the group to complete the analysis report (days)
2015	306	19
2016	329	16
2017	334	5
2018	332	3
2019	341	2
2020	356	2

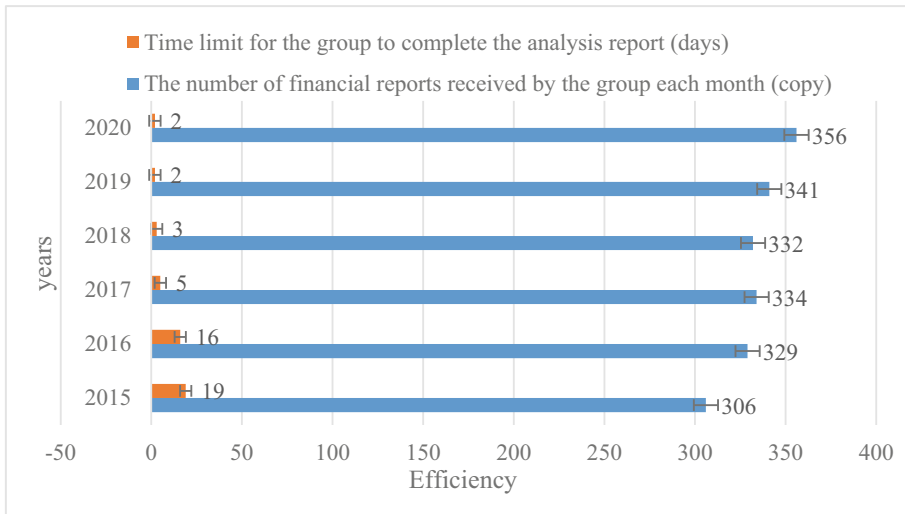


Fig. 2. Financial analysis efficiency

Looking at Fig. 2, we can find that the number of financial reports received by the group has increased significantly after the system was launched, but the time required for the group to complete the analysis report each month has decreased significantly. This shows that the application of “blockchain + 5G” technology to the financial system not only greatly reduces the workload of financial analysis, but also solves the problem of information asymmetry, which is convenient for the group to grasp the financial affairs of its subsidiaries in a timely, convenient, comprehensive and accurate manner.

5 Conclusions

“Blockchain + 5G” technology, as an innovative technology representative in the era of rapid Internet development, has now played its irreplaceable role in the financial field by virtue of its own characteristics and advantages and has continuously triggered changes. This article combines the “blockchain + 5G” technology with finance, and explores the innovative application of this technology to the digital intelligence of finance. This article analyzes the “blockchain + 5G” technology environment, which can promote the intelligent storage of financial statement data, simplify the supervision and audit process, reduce supervision and audit costs, improve the independence of internal audit, and promote the management of capital structure. Therefore, the development of finance to digital intelligence is inseparable from the technical guarantee of blockchain and 5G technology.

Acknowledgements. This work was supported by the Quality Edition Promotion Project of Zhanjiang Science and Technology College: Data Intellectualized Finance Teaching team. (project No.:PPJH202107JXTD).

References

1. Du, M., Chen, Q., Xiao, J., et al.: Supply chain finance innovation using blockchain. *IEEE Trans. Eng. Manage.* **PP**(99), 1–14 (2020)
2. Wang, R.L., Wu, Y.: Application of blockchain technology in supply chain finance of Beibu gulf region. *Math. Probl. Eng.* **2021**(1), 1–10 (2021)
3. Chen, J., Cai, T., He, W., et al.: A blockchain-driven supply chain finance application for auto retail industry. *Entropy* **22**(1), 95 (2020)
4. Chen, A.Q.: Studies on the effect of Beijing-Tianjin-Hebei collaborative development on the digitization of inclusive finance in the region along the route. *Finance* **11**(3), 176–187 (2021)
5. Manuela, V., Tucker, B., Daniel, B., et al.: Artificial intelligence research in finance: discussion and examples. *Oxf. Rev. Econ. Policy* **37**(3), 3 (2021)
6. Desbiolles, J.P.: Finance et Intelligence artificielle (IA): d’une révolution industrielle à une révolution humaine. tout est à repenser. *Annales des Mines - Réalités industrielles* **2019**(1), 5 (2019)
7. Lee, J.: Access to finance for artificial intelligence regulation in the financial services industry. *Eur. Bus. Organ. Law Rev.* **21**(4), 731–757 (2020). <https://doi.org/10.1007/s40804-020-00200-0>
8. Begenau, J., FaRboodi, M., Veldkamp, L.: Big Data in Finance and the Growth of Large Firms. *J. Monetary Econ.* **97**(AUG.), 71–87 (2018)
9. Wang, L., Luo, X.R., Lee, F., et al.: Value creation in blockchain-driven supply chain finance value creation in blockchain-driven supply chain finance. *Inf. Manage.* 2021(3), 103510 (2021)
10. Kaufmann, S., Frühwirth-Schnatter, S., Dijk, H.: Editorial introduction on complexity and big data ineconomics and finance: recent developments from a Bayesian perspective. *J. Econ.* **210**(1), 1–3 (2019)
11. Westland, J.C.: Introduction to the special issue on big data in finance and business. *Electron. Commer. Res.* **18**(2), 201–201 (2018). <https://doi.org/10.1007/s10660-018-9306->
12. Begenau, J., Farboodi, M., Veldkamp, L.: Big data in finance and the growth of large firms. *J. Monetary Econ.* **97**(AUG.), 88–90 (2018)



Comprehensive Evaluation Model of TOPSIS Method Based on Data Processing Technology

Yi Huang¹ and Wenbin Liu^{1,2}(✉)

¹ Wuhan Technology and Business University, Wuhan, Hubei, China

² Wuchang Institute of Technology, Wuhan, Hubei, China

Abstract. Compare with the traditional multivariate statistical methods of data processing technology for evaluating problems, TOPSIS method has the characteristics of intuitive analysis principle, simple calculation and less requirement on sample size. Therefore, this paper adopts entropy weight TOPSIS method to establish a comprehensive evaluation model to accurately reflect the gap between suppliers. And extracts the supplier's market share, the stability of supply and demand relationship, supplier reputation as evaluation indexes by mining historical data, establishes a comprehensive evaluation model of supplier's supply capacity, and carries on the quantitative analysis of the supply characteristics. Based on the average supply quantity, transaction times and the probability of supply quantity exceeding or not reaching the enterprise's order quantity in the past five years, the weight of each evaluation index is obtained by using entropy weight method. And the evaluation score that can reflect the importance of production is obtained by MATLAB. Finally, 10 suppliers are identified as the most important suppliers.

Keywords: Entropy weight method · TOPSIS method · Data processing · Comprehensive evaluation model · MATLAB

1 Introduction

In the context of modernization, the supply, transportation, storage, production and other links of raw materials in large-scale production enterprises, which can become the focus of competition of enterprises in the market. Each link is closely linked. Appropriate distribution and transportation scheme has very important strategic significance for the operation and development of production enterprises. With the rapid development of big data era, production and transportation technology is increasingly mature. The order and transport of raw materials, the allocation of inventory and capacity, which coordinate communication with raw material suppliers, third-party logistics companies and production enterprises. Construct the economical plan of cost and loss of raw material production line. The application of appropriate distribution scheme to production line can provide information support for the optimization control and decision-making of raw material logistics process. Generate optimal raw material guarantee plan and scheme to provide scientific and reasonable auxiliary decision-making information for decision makers.

2 Data Preprocessing

The raw materials used by a building and decorative board manufacturer, are mainly composed of wood fiber and plant fiber, which can be divided into three types: A, B and C. The enterprise needs to make the raw material ordering plan in advance every year, that is, determine the raw material suppliers to be ordered according to the capacity requirements.

The partial data of order quantity and supply quantity of 402 raw material suppliers in the recent 5 years (240 weeks) are shown in Tables 1 and 2 respectively.

Table 1. Supply quantity of supplier

Supplier ID	Material classification	W001	W002	W003	W004	W240
S001	B	0	0	0	0	0
S002	A	1	0	0	1	1
S003	C	8	1	0	0	11
S004	B	0	0	0	0	0
S005	A	37	62	60	65	81
S006	C	0	0	0	0	0
S007	A	9	94	37	102	17
S008	C	0	0	0	0	0
.....
S402	B	0	0	0	0	0

Table 2. Order quantity of enterprise

Supplier ID	Material classification	W001	W002	W003	W004	W240
S001	B	0	0	0	43	0
S002	A	1	1	0	1	1
S003	C	7	1	0	0	10
S004	B	0	1	1	100	0
S005	A	30	60	60	60	80
S006	C	0	0	0	0	0
S007	A	8	150	30	200	10
S008	C	0	0	0	0	1
.....
S402	B	0	0	0	0	0

By looking at 240 weeks of historical data, if there are more than 5 similar outliers, the supplier is considered capable of delivering these quantities. If there are less than 5 similar outliers, these values are considered as outliers. Deleting these data has little impact on the sample, so it is directly deleted as seen in Table 3.

Table 3. Weeks of abnormal

Supplier ID	Weeks of abnormal
S151	W83,W84,W85
S201	W85
S348	W138,W139,W140
S374	W237,W238,W239,W240

3 Quantitative Analysis of Supply Characteristics Based on Comprehensive Evaluation Model

3.1 The Evaluation Index

Market share (x_1): refers to the share of a certain product in the market [1, 2], which can reflect the position of the enterprise in the market. In this paper, market share refers to the supply quantity of each supplier to the enterprise, which can reflect the supplier’s supply influence on the enterprise. Here, the average of all supply quantities of 240 weeks is used to reflect the market share, denoted as x_1 . The larger the value is, the higher the market share is.

Supplier reputation (x_2): As can be seen from the data in Tables 1 and 2, the order quantity of the enterprise and the supply quantity of the supplier are often not equal, and the actual supply may be more or less than the order quantity. For this reason, the probability that the supply quantity of each supplier is greater than or less than the order quantity is calculated respectively. A more detailed indicator reflecting the supplier’s reputation is introduced. The probability that the supply quantity is greater than the order quantity is denoted as indicator x_{21} . The probability that the supply quantity is less than the order quantity is denoted as indicator x_{22} .

Stability of supply and demand relationship (x_3): For each supplier, the enterprise does not generate orders for it every week, that is, it does not generate supply and demand relationship every week. Here, the number of weeks generating supply and demand relationship within 240 weeks is calculated, denoted as x_3 . The higher the value, the more stable the supply and demand relationship.

3.2 Determine the Weight of Each Indicator

Based on the above three indicators, it is necessary to judge the influence of each indicator on the supply characteristics. Entropy value [3, 4] can judge the influence of multiple indicators on comprehensive evaluation, so the entropy weight method is used to calculate the weight of each indicator.

In general, in comprehensive evaluation, the greater the variation degree of index value, the smaller the information entropy, the more information the index provides, and the larger the weight should be. Therefore, the weight of each index by using entropy – entropy weight, can be calculated according to the variation degree of each index value [5, 6].

Due to the dimension difference of indicators, in order to eliminate the influence of dimension, each indicator is normalized first. The standardized matrix X of 402 suppliers' data and 3 indicators was obtained:

$$X_{ij} = \frac{x_{ij} - \min(x_{kj})}{\max(x_{kj}) - \min(x_{kj})} \quad (1)$$

Where $i, k = 1, 2, \dots, 402$ represents 402 suppliers and $j = 1, 2, 3$ represents 3 evaluation indicators respectively.

Calculate the proportion p_{ij} ($i = 1, 2, \dots, 402; j = 1, 2, 3$) of the i supplier's data in the j index, and regard it as the probability to be used in the relative entropy operation, obtain the probability matrix P , and the calculation formula is as follows:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{402} x_{ij}} \quad (2)$$

The information entropy of the index j :

$$e_j = -\frac{1}{\ln 402} \sum_{i=1}^{402} p_{ij} \ln(p_{ij}) \quad (3)$$

Calculate information utility value d_j , and the entropy weight of each index W_j is obtained by normalization:

$$d_j = 1 - e_j \quad (4)$$

$$W_j = \frac{d_j}{\sum_{j=1}^3 d_j} \quad (5)$$

According to the above method, market share (x_1), reputation of supply quantity greater than order quantity (x_{21}), reputation of supply quantity less than order quantity (x_{22}) and stability of supply and demand relationship (x_3), can be obtained as Table 4:

Table 4. The weight of evaluation indicators

Indicator	x_1	x_{21}	x_{22}	x_3
Weight	0.17868	0.28857	0.25996	0.27279

It can be seen from the Table 4 that, the reputation of the supply quantity exceeding the order quantity has the greatest influence on the supplier’s supply characteristics, while the market share has the least influence.

3.3 Comprehensive Evaluation Model Based on TOPSIS Method

Compare with the traditional multivariate statistical methods for evaluating problems, TOPSIS method has the characteristics of intuitive analysis principle, simple calculation and less requirement on sample size. Therefore, this paper adopts entropy weight TOPSIS method [7, 8] to establish a comprehensive evaluation model of supplier’s supply capacity.

According to the normalization matrix X obtained from the above forward transformation and standardization of 402 suppliers and 3 indicators, new definitions are given:

The maximum value:

$$\begin{aligned}
 X^+ &= (X_1^+, X_2^+, X_3^+) \\
 &= (\max\{X_{11}, \dots, X_{i1}\}, \dots, \max\{X_{13}, \dots, X_{i3}\})
 \end{aligned}
 \tag{6}$$

The minimum value:

$$\begin{aligned}
 X^- &= (X_1^-, X_2^-, X_3^-) \\
 &= (\min\{X_{11}, \dots, X_{i1}\}, \dots, \min\{X_{13}, \dots, X_{i3}\})
 \end{aligned}
 \tag{7}$$

Distance between each indicator of the supplier and the maximum value:

$$D_i^+ = \sqrt{\sum_{j=1}^3 W_j \cdot (X_j^+ - x_{ij})^2}
 \tag{8}$$

Distance between each indicator of the supplier and the minimum value

$$D_i^- = \sqrt{\sum_{j=1}^3 W_j \cdot (X_j^- - x_{ij})^2}
 \tag{9}$$

So far, the comprehensive evaluation model based on TOPSIS method has been established.

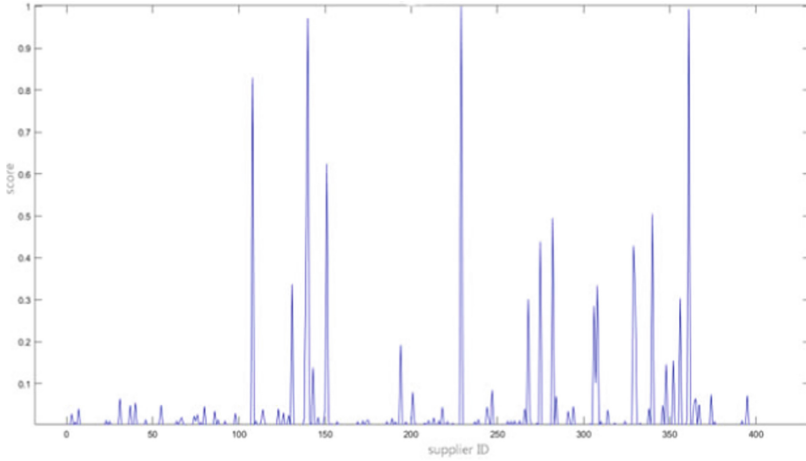


Fig. 1. Score line chart of 402 suppliers' supply characteristics

3.4 Model Solving by MATLAB

MATLAB is used to solve the above comprehensive evaluation model [9, 10], and the supply characteristics score line graph of 402 suppliers is obtained as Fig. 1.

As can be seen from the line chart, the scores of each supplier are not the same, and there are large differences. In the comprehensive evaluation model, the higher the score is, the higher the production guarantee is for the enterprise. According to the score to sort, it can get the order of 402 suppliers score. The supplier with the highest score is S229, whose material classification is A, and score is 1.

Select the top 10 suppliers in the ranking, which are the most important suppliers to ensure the production of enterprises. There are 4 suppliers whose material classification is A, 4 suppliers whose material classification is B and 2 suppliers whose material classification is C. Their respective ID and scores are shown in the Table 5:

Table 5. Top 10 most important suppliers and scores

Supplier ID	Material classification	Score	Supplier ID	Material classification	Score
S229	A	1.0000	S340	B	0.5055
S361	C	0.9937	S282	A	0.4946
S140	B	0.9712	S275	A	0.4389
S108	B	0.8293	S329	A	0.4286
S151	C	0.6248	S139	B	0.3987

4 Conclusions

With the increasing maturity of production and transportation technology, the supply, transportation, storage and other links of raw materials of large-scale production enterprises are the focus of enterprise competition. Therefore, the study of appropriate ordering scheme is of great significance to improve the operation quality of production enterprises.

In order to quantitatively analyze the supply characteristics of the suppliers, the market share, stability of supply and demand relationship, and reputation ability are extracted as the supply characteristics according to the data of order quantity and supply quantity in the past 240 weeks. Then establish a comprehensive evaluation model based on the entropy weight TOPSIS method, to evaluate the importance of production of each supplier. Finally, the 10 most important suppliers are selected from the highest to the lowest rating.

References

1. Beaumont, E.: Market share. *Archit. Rev.* **242**(1447), 20–27 (2018)
2. Palmer, K., Tate, J.E., Wadud, Z., Nellthorp, J.: Total cost of ownership and market share for hybrid and electric vehicles in the UK US and Japan. *Appl. Energy* **209**(1), 108–119 (2018)
3. Zhang, L., Li, S., Zheng, D., Shi, S., Zeng, P.: Evaluation of thermal recovery heavy oil development effect based on interval entropy TopSIS. *J. Xi 'an Shiyu Univ. Nat. Sci. Ed.* **35**(2), 6–10 (2020). (in Chinese)
4. Rao, K., Yang, Z.: Evaluation of mountain land ecological security based on Dpsir model and entropy weight method. *Asian Agric. Res.* **12**(02), 32–37 + 41 (2020)
5. Dastbaz, M., Arabnia, H., Akhgar, B.: Evaluation of the dynamic cybersecurity risk using the entropy weight method. In: *Technology for Smart Futures*, pp. 271–287. Springer, Cham (2018). <https://doi.org/10.1007/978-3-319-60137-3>
6. Zhou, H., Wang, Y., Zhang, M., Zhang, H., Zhang, H., Wang, L.: Fuzzy comprehensive evaluation of secondary equipment efficiency in smart substation weighted by entropy method. *Electr. Measur. Instrum.* **55**(3), 7–8 (2018) (in Chinese)
7. Mohammed, M.A., Abdulkareem, K.H., Al-Waisy, A.S., Mostafa, S., Díez, I.D.L.T.: Benchmarking methodology for selection of optimal covid-19 diagnostic model based on entropy and topsis methods. *IEEE Access* **8**(1), 99115–99131 (2020)
8. Seker, S., Aydin, N.: Hydrogen production facility location selection for black sea using entropy based topsis under Ivpf environment–sciencedirect. *Int. J. Hydrogen Energy* **45**(32), 15855–15868 (2020)
9. Han, Z.: *Mathematical Modeling Method and Application*. Higher Education Press, Beijing (2017) (in Chinese)
10. Si, S., Sun, X.: *Mathematical Modeling Algorithm and Application*. National Defense Industry Press, Beijing (2021) (in Chinese)



Data Mining Algorithm of Hotel Customer Data Based on Mobile Computing

Jun Wang^(✉)

Shenyang Polytechnic College, Shenyang 110045, Liaoning, China

Abstract. Mobile computing has gradually emerged and developed vigorously along with the development of computer technology, Internet communications and networks. As a new and practical and efficient technical means, mobile computing has outstanding advantages such as strong operability, wide range of use, high information processing capabilities and fast calculation speed, and has been widely used in all aspects of social life. At the same time, the hotel industry, as an important part of the service industry, needs to analyze user needs and provide personalized customized services, so there is a greater demand for hotel customer data mining. In the mobile computing environment, the research of customer information data mining has become one of the indispensable and important contents of today's social development. This article adopts experimental analysis and data analysis methods to understand the effect of clustering analysis of different data mining algorithms through experimental research, so as to improve the ability of the algorithm to analyze customer information. According to the experimental results, among these algorithms, the clustering time of the FDBSCAN algorithm is the shortest, which can achieve the clustering effect of the DBSCAN algorithm, and the clustering time and time complexity are both smaller than the DBSCAN algorithm, and it takes advantage of reducing the time complexity.

Keywords: Mobile computing · Customer information · Data mining · Cluster analysis

1 Introduction

As the development trend of the combination of the hotel industry and Internet information technology becomes increasingly obvious, the use of mobile data mining algorithms to analyze user needs and personalize customized services is the direction of future hotel industry research. The users in the hotel information found some potentially useful and valuable content during data mining. Applying data mining algorithms to the analysis of hotel customer data, under the premise of ensuring privacy and security, is conducive to bringing more personalized services to customers, and at the same time, it also helps hotels and other enterprises to make decisions and improve work and service quality.

At present, some scholars have conducted research on mobile computing and customer information data mining. For example, Lalicic L pointed out that Mobile computing platforms allow consumers to creatively create content and share it with others,

which is certainly an innovation [1]. Han Qin believes that the data mining model is highly predictive and can effectively predict telecom customers with high churn rate, so as to formulate specific and effective customer retention programs to reduce their churn rate [2]. Wang Q proposed that Mobile user equipment can be processed quickly through the technical means of mobile edge computing, so as to provide users with an excellent experience [3]. Therefore, this article starts from a new perspective, combined with mobile computing, and conducts research on hotel customer data mining algorithms, which has important research significance and value.

This article mainly discusses these aspects. First, the concept of mobile computing and related theoretical knowledge are explained. Then, it discusses the content of customer information data mining. In addition, it also introduces commonly used data mining algorithms for customer acquisition. Finally, the experimental research is carried out around the clustering effect of the data mining method, and the relevant experimental results and analysis conclusions are drawn.

2 Related Theoretical Overview and Research

2.1 Mobile Computing

Mobile computing has been widely used in various fields in today's society. It can provide users with a variety of services, such as various navigation, wireless network communications, and data transmission. With the development and popularization of smart phones and the application of 5G technology, mobile computing has become an inevitable trend and demand. For example, mobile phones are the most important and commonly used mobile terminal devices in people's lives. In addition, there are mobile terminal devices such as tablets or WIFI.

Mobile computing is a new computer application based on the development of Internet technology. It can realize rapid calculation, automatic storage and management when providing various services to users. In essence, it is to complete data information exchange and sharing through terminal devices such as wireless networks or mobile phones. At present, many hotels have also begun to adopt this method, and have effectively used their own resources and achieved certain results. However, for most people, they are still in their infancy or in the early stage of research. Mobile computing technology It is not yet mature and standardized [4, 5].

Mobile computing is a part of computer applications, and it is also one of the most basic, most valuable and most concerned technologies in the field of mobile communications. In the era of big data, people are increasingly dependent on information processing. With the development of technology and the continuous update of user needs, these different types and functional requirements have put forward higher requirements for us, how to quickly and accurately obtain the required data from the massive complex, multi-source heterogeneous format.

The mobile computing system is composed of multiple independent subnets, which realize information exchange and resource sharing through data transmission and processing. It includes mobile communication technology and network technology. People

can choose a suitable calculation method according to different application environments and apply it to specific problems. At the same time, it also takes into account the influence of wireless channels on signal quality [6, 7].

In the data transmission process of mobile computing, it is necessary to consider the ability and efficiency of information processing in different environments. At present, there are many softwares that can realize various complex data transmission, but these are all based on specific scenarios for analysis and research. Therefore, people's needs for high speed, large throughput, etc. cannot be met under the background of the current era. In addition, due to the huge and unstable data volume, its application is limited. The general flow of mobile computing is shown in Fig. 1.

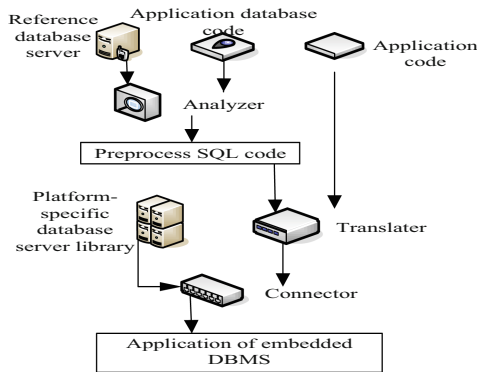


Fig. 1. The general process of mobile computing

2.2 Customer Profile Data Mining Algorithm

Today, customer information is an important resource. How to effectively use these data has become one of the hottest topics now. After long-term accumulation, companies have collected and stored a large amount of valuable customer data, which will provide a basis for decision-making for hotel development.

Customer information management system is an important part of enterprise information management. It can help managers analyze scientifically and reasonably, predict and control the risks they will face in business activities. At the same time, it also provides a more comprehensive, systematic, practical and convenient for decision makers.

With the widespread use of data mining tools, telecommunications and computer networks, protecting the privacy and information security of companies or individuals has become a major problem in data mining. People should ensure the confidentiality and security of information while ensuring the normal access to information [8, 9].

The use of various data analysis and extraction techniques may lead to the destruction of privacy. For example, the release of medical records in which the user's identity is hidden when used for analysis may also result in the loss of medical information. Browsing privacy-protected data provides a way to solve this problem. After the data is

preprocessed, meaningful models can still be extracted without seriously affecting the usability of the data, and the privacy information implicit in the original data is protected.

Therefore, in the follow-up work, it is necessary to consider the problem of data mining leaking hotel customer information, and overcome the shortcomings of generalization technology, so as to improve the strength of privacy protection, and strengthen the security and the availability of anonymized data.

The data mining method refers to the analysis and classification of a large amount of original information, and then discovers the underlying laws, and studies the objects of different types of individuals or groups that have the same attributes and different types of characteristics as a whole.

Under uncertain conditions, classify a large amount of original data, extract useful information and convert it into technical means that can provide applications. This method is also called unsupervised learning algorithm or self-organizing learning method. It can obtain relevant knowledge points such as the attributes of the data object and its inherent regular characteristics from different angles according to the needs to achieve the purpose of prediction, and adopt a combination of multiple methods for processing.

First of all, according to different needs, choose what you need to obtain knowledge or information content related feature value minimization, model parameter estimation problem. Then, perform classification processing in an uncertain environment. Divide the original data into several parts for predictive research and analysis, and then draw conclusions and propose targeted measures to improve decision-making efficiency and effects. A large number of potential contradictions and problems that cannot be solved by existing methods should be considered from multiple angles. In order to establish a new decision-making process or mode [10, 11].

Data mining is based on a large amount of customer information and forms new opinions and things by sorting out and analyzing the behavioral laws of massive hotel users and other potentially useful or useless but unobtainable knowledge. In the process of data processing, the valuable and predictive information needs to be extracted and used.

Choosing the right data mining algorithm is the core of obtaining customer information. The choice of algorithm must consider the type of customer data and mining tasks. For example, the data mining algorithms commonly used for customer acquisition include the K-Means algorithm, the DBSCAN algorithm, and the ID3 algorithm.

The ID3 algorithm is simple and easy to use, and the results of the algorithm are easy for users to understand, and there is no need to understand a lot of basic knowledge during the learning process. As long as the training examples can be individually displayed as attributes, the algorithm can be used for training. The algorithm is considered to be simple, efficient and reliable, and its position in inductive learning is still very important.

In addition, the K-Means algorithm has the advantages of reliable operation, simple, fast convergence, and efficient processing of large data sets, which has been widely used. Iteration makes the selected cluster center closer and closer to the true center of gravity of the cluster. And the clustering effect is getting better and better. The algorithm still has some shortcomings, and it needs to be continuously adapted to the classification of

samples [12]. The specific calculation method is shown in formula (1, 2).

$$K = \sum_{u=1}^l \sum_{k=1}^{m_u} e_{uk}(y_k, f_u) \quad (1)$$

$$K_f = \sum_{k=1}^l \sum_{l=1}^{m_u} \left| y_l^{(k)} - f_k(U) \right|^2 \quad (2)$$

Among them, K is the objective function, which is used to judge whether the clustering effect is good or bad. The smaller the value of K , the tighter and more independent the clusters. l is the number of cluster centers in the data set, e is the dimension of the data set, and m_u is the number of data points in the cluster. e_{uk} is a measure between y_k and f_u . K_f is a criterion function. After it converges, it means that the operation of the K-means algorithm is over.

3 Experiment and Research

3.1 Experimental Background

Cluster analysis is an important data mining method. Among many clustering methods, the DBSCAN algorithm is a classic density-based clustering algorithm. However, no one method of data mining can be applied to all applications. Therefore, it is necessary to further optimize this type of clustering analysis algorithm to adapt it to different application requirements.

3.2 Experimental Environment

The software environment of this experiment is Eclipse SDK 3.4.7, Windows XP operating system. The hardware environment is a PC with a CPU Intel Core 2 Duo T7100 (2.0 GHz) and 4 GB memory.

3.3 Experimental Process

This experiment aims to reduce and improve the runtime complexity of data mining algorithms, and design a fast clustering algorithm FDBSCAN algorithm based on customer information. In order to understand the clustering analysis effect of the algorithm more intuitively, this experiment selected more common algorithms, including DBSCAN algorithm, PDBSCAN algorithm and SA-DBSCAN algorithm, and tested the clustering time of these algorithms. Some experimental results are shown below.

4 Analysis and Discussion

In this experiment, four data sets were selected as experimental samples to test the clustering time of FDBSCAN, DBSCAN, PDBSCAN and SA-DBSCAN algorithms. The test results are shown in Table 1.

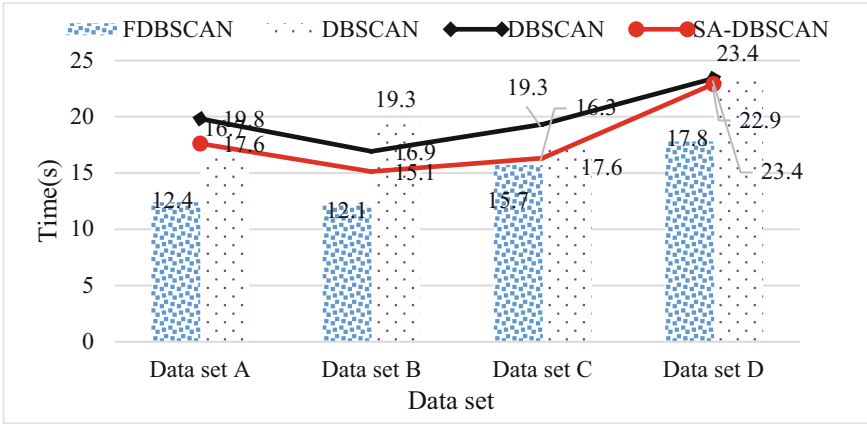


Fig. 2. Test results of algorithm running time

Table 1. Test results of algorithm running Ttime

Data set	FDBSCAN	DBSCAN	DBSCAN	SA-DBSCAN
Data set A	12.4	16.7	19.8	17.6
Data set B	12.1	19.3	16.9	15.1
Data set C	15.7	17.6	19.3	16.3
Data set D	17.8	23.4	23.4	22.9

It can be seen from Fig. 2 that the clustering time of the FDBSCAN, DBSCAN, PDBSCAN and SA-DBSCAN algorithms are 12.4, 16.7, 19.8 and 17.6s respectively for the same data set, and the clustering time of the FDBSCAN algorithm is the shortest among these algorithms. It can be seen that the FDBSCAN algorithm can achieve the clustering effect of the DBSCAN algorithm, and the clustering time and time complexity are less than the DBSCAN algorithm, and it has a greater advantage in reducing the time complexity.

5 Conclusion

With the continuous improvement of people’s living standards, the requirements for service industries such as hotels are also getting higher and higher. Therefore, data mining must be used to discover potential customer groups. The era of mobile computing has brought huge opportunities for the development of hotels, and statistical analysis and prediction of hotel customer data has become particularly important. Applying data mining algorithms to the analysis of hotel customer data, under the premise of ensuring privacy and security, is conducive to bringing more personalized services to customers, and at the same time, it also helps hotels and other enterprises to make decisions and improve work and service quality. This paper is based on mobile computing

and researches on hotel customer data mining algorithms, which conforms to the trend and trend of the development of the times, and has important research significance and value.

References

1. Lalicic, L., Dickinger, A.: An assessment of user-driven innovativeness in a mobile computing travel platform. *Technol. Forecast. Soc. Chang.* **144**(JUL.), 233–241 (2019)
2. Qin, H.: Forecast analysis of churn of telecom customers based on data mining algorithm. *Fujian Qual. Manag.* **000**(024), 216–217 (2018)
3. Wang, Q., Gao, A., Hu, Y.: Joint power and QoE optimization scheme for multi-UAV assisted offloading in mobile computing. *IEEE Access* **PP**(99), 1–1 (2021)
4. Singh, S.K., Vidyarthi, D.P.: A heuristic channel allocation model with multi lending in mobile computing network. *Int. J. Wireless Mobile Comput.* **16**(4), 322–339 (2019)
5. Yin, L., Guo, Y., Zhang, H., et al.: Threat-based declassification and endorsement for mobile computing. *Chin. J. Electron.* **28**(005), 1041–1052 (2019)
6. Li, Y., Kim, K., Ding, Y.: Early warning system of tennis sports injury risk based on mobile computing. *Mob. Inf. Syst.* **2021**(3), 1–10 (2021)
7. Zhang, S., Mao, H.: Optimization analysis of tennis players' physical fitness index based on data mining and mobile computing. *Wireless Commun. Mob. Comput.* **2021**(11), 1–11 (2021)
8. Liu, S., Xu, C., Xiong, X., et al.: Dynamic scheduling model of rail-guided vehicle (RGV) based on genetic algorithms in the context of mobile computing. *Int. J. Mob. Comput. Multimedia Commun.* **12**(1), 43–62 (2021)
9. Alkhalil, A.: Evolution of existing software to mobile computing platforms: Framework support and case study. *Int. J. Adv. Appl. Sci.* **8**(3), 100–111 (2021)
10. Yan, J.Z., Li, J.E., Li, Q.L.: A study on the personalization recommendation service incorporating review-based customer profile. *J. Digit. Contents Soc.* **21**(9), 1575–1584 (2020)
11. Abreu, T., Minussi, C.R., Lopes, M., et al.: Electrical customer profile using fuzzy logic theory. *IEEE Lat. Am. Trans.* **18**(8), 1353–1361 (2020)
12. Mohadab, M.E., Bouikhalene, B., Safi, S.: Automatic CV processing for scientific research using data mining algorithm. *J. King Saud Univ. Comput. Inf. Sci.* **32**(5), 561–567 (2020)



Design and Development of Interior Design CAD System Based on Data Mining Algorithm

Yun Peng¹ (✉) and Feilu Hang²

¹ College of Liberal Arts, Yunnan Technology and Business University,
Kunming, Yunnan, China

² Information Center of Yunnan Power Grid Co., Ltd, Kunming, China

Abstract. In the field of interior design, the relationship between science and art has always been the focus of attention. With the rapid development of science and technology today, the influence of computer technology in the field of interior design continues to deepen, and it brings new visual experience and work methods to interior designers. This article aims to study the design and development of the interior design CAD system based on data mining algorithms. Based on the analysis of the advantages of data mining methods, decision tree algorithms and computer-aided design, the interior design CAD system is designed, and the performance of the system and user satisfaction were tested and investigated. The results show that the response time of the system is within 3 s, which meets the requirements of this article, and for users, the system of this article has strong practicability.

Keywords: Data mining algorithm · Interior design · CAD system · System design

1 Introduction

Due to the rapid development of modern computer technology, computers have had a huge impact on people's work, lifestyle, and ways of thinking. The various environments in which humans live have become more and more extensive and are covered by computers and networks. The relationship between human brain and computer is also undergoing tremendous changes [1, 2]. It is in this context that computers have become a major part of today's architectural design industry. In modern interior design, both interior design methods and content design require designers to seek new interior design strategies and methods that conform to the changes in the law of modern science and technology [3, 4].

CAD is used to implement CAD technology on a computer. In short, it is a 2D design software with powerful drawing, editing, patterning, size adjustment and hatching functions [5]. There are also some 3D drawing tools, but compared with commonly used 3D drawing software, the degree of simulation is very low. Thanks to these practical functions, CAD software is widely used in various fields, especially in the fields of architecture and interior design [6, 7]. It can be used for preliminary interior design, 3D

modeling, architectural performance design, community design, and interior design. It also has the characteristics of easy to learn, easy to use, simple to operate, powerful and open structure. It is not only convenient for users to use, but also can be used by customers for secondary development. The software can be continuously expanded and improved. Therefore, AutoCAD is widely used Microcomputers and various workstations [8, 9].

On the basis of consulting a large number of relevant references, this paper combines the common methods of data mining, decision tree algorithms and the advantages of computer-aided design, design an interior design CAD system, and test the performance of the system and user satisfaction. And investigate to verify whether the system meets the requirements.

2 Design and Development of Interior Design CAD System Based on Data Mining Algorithm

2.1 Common Methods of Data Mining

(1) Classification

The classification method mainly refers to finding the common characteristics of one or more data objects in the database system, and classifying them according to a certain classification mode [10, 11]. The main purpose is to assign certain data elements in the database system to a certain type. This can be used in customer types, customer characteristics and potential, customer satisfaction, and prediction of customer market trends. For example, new car sellers can divide them into different types according to customers' preference for car models. Based on these preferences, various categories of new car brochures can be mailed to users, which can significantly expand business opportunities.

(2) Clustering

The method of cluster analysis is to divide the data set into several types through the difference or similarity of the data types. Its main purpose is to minimize the similarity between different types of data, thereby reducing the similarity between the same type of data.

(3) Association rules

Association rule learning is a rule that describes the relationship between data elements in the database. In other words, based on the appearance of certain elements in the transaction, other elements may appear in the same transaction. That is, the hidden relationships in the data or the relationships between them.

(4) Features

The feature analysis method is to extract the attribute type of the data from the data set. These attributes represent the overall characteristics of the data set. For example, by deriving the main characteristics of customer rejection, merchants can obtain these main reasons and characteristics of customer rejection, and then avoid a large amount of customer loss based on these reasons and characteristics.

(5) Analysis of changes and deviations

Bias includes a large number of possible and interesting insights, such as classification anomalies, abnormal patterns, and expected observation biases. Unexpected rules can bring huge benefits. Once the potential value of these abnormal rules is discovered, the benefits will be immeasurable.

2.2 Decision Tree Algorithm

(1) ID3 algorithm

ID3 algorithm is a process of constructing and analyzing decision data based on information gain and entropy. The definition of information gain is positively related to the probability of the data. Algorithm ID3, when defining the attributes of each node in the decision tree, uses information gain to define the attribute pattern and generate the attribute information size of the attribute to define the final decision tree pattern. Overall implementation, this process is based on the requirement for continuous traceability completion time.

Since the purpose of data mining is to collect uncertain, random, non-uniform, and complex data, it is necessary to describe the dynamic characteristics of the data while quantifying the data information. The ID3 algorithm uses conceptual entropy to determine this index. This represents an indeterminate process of notifying data. Normally, a data value of 0 indicates that the data is certain, and a data value of 1 indicates that the data is uncertain. The normal entropy value is set between 0 and 1. If all data subsets show the same probability of related data, it means that the entropy is the highest and the data is the most stable.

The probability of a given set of data is p_1, p_2, \dots, p_n , where the condition is met:

$\sum_{i=1}^n p_i = 1$ defines the value of entropy $H(p_1, p_2, \dots, p_n)$ as:

$$H(p_1, p_2, \dots, p_n) = \sum_{i=1}^n (p_i \log(1/p_i)) \quad (1)$$

$$Gain(D, M) = H(D) - \sum_{i=1}^n p(m_i)H(m_i) \quad (2)$$

(2) C4.5 algorithm

On the basis of the ID3 algorithm, the algorithm has been appropriately improved. The main improvements include redefining the defining characteristics of information gain and the discrete transformation of continuous data. Although the implementation process of this algorithm can achieve more stable and efficient decision tree construction and pruning, it also has some shortcomings. The main reason is that its algorithm execution efficiency is low. The efficiency of time and space complexity determines that it cannot meet the requirements of computer program execution processing. The need for large amounts of data.

(3) SLIQ algorithm

The SLIQ algorithm is further refined on the basis of algorithm C4.5, and applies the idea of norm and amplitude algorithm to the classification and execution process of decision numbers. Among them, the standard idea is to classify the relevant target data set when creating the decision number, which is the same as the previous isolation standard for each attribute. The basis of classification is to increase the order of information acquisition to achieve the first range of execution efficiency ideas. Although the SLIQ algorithm achieves the goal of optimizing efficiency, the time complexity of the algorithm is still nonlinear time, and there is a performance bottleneck in processing large amounts of data.

2.3 Advantages of Computer Aided Design

(1) Provide a very large free space for interior designers

Designers can use the computer to collage, combine the drawing materials that the computer can master, and make various attempts in a simple and convenient way to create works that are most suitable for the current architectural environment. This convenient experimental method helps to inspire the creators.

(2) Make the interior design more accurate

Errors in the manual design process are inevitable, and computers can make up for this shortcoming. With the help of computer software, many designs can be made more accurate, and the design results can be identified in advance through the preview function to avoid unnecessary problems such as duplication and error failure. In addition, computer-aided design is more realistic than traditional manuscripts, allowing people to see the actual shape, color, texture, environment and other final effects of future interior design through the screen. Traditional hand-painted designs want to minimize the gap between the design and the actual scene through various design drawings and manual models, but there are also some unavoidable factors. For example: when designing a perspective design, we want to use realistic visual effects to check whether the design is meaningful. However, most designers of traditional manual perspective design tend to ignore or deliberately dilute the influence of colors and materials, and do not pay enough attention to the influence of light factors in the room, resulting in very low design effects and not real. The introduction of computers can accurately solve these problems. The introduction of computers for interior decoration is very useful to designers and also very helpful to users.

(3) Simple and fast

Interior designers have a deep understanding of this advantage. For example, if you choose an error in the design process, you can choose to cancel the function immediately, but manual design cannot do this, which can save a lot of time and improve design efficiency. It saves people's time greatly in duplication and modification, and the computer-assisted system improves work efficiency. For example, if you are designing 6 identical dining chairs and you are using a traditional manual design, you need to design them one by one. However, if you can use the power of computer technology, you can use the same commands to perform this function. In the field of interior design, designers can flexibly use commands such as mass production and quick cancellation to improve work efficiency. At the same time, when the computer system creates a computer model, it can change the perspective to obtain different side models, and appropriately decompose each part of the model to view the structure of the building.

(4) Easy to carry

Interior designers can easily carry works or semi-finished products in the company and at home, and do simultaneous updates and perfect operations, without the need for heavy tools like manual design. At the same time, designers can easily share their work with others to receive suggestions from others and sell them to others, thus saving time and space. Computer-aided design can save blueprints at any time, call them at any time, and save drawings at various stages of the design process. These functions allow designers to call design plans at different stages for customization, comparison and modification at any time, which plays an important role in the gradual improvement of the design.

3 Experiment

3.1 System Login Module

The login system use case is mainly to verify the user's login. When the user enters the registered user name and password, the system verifies its legitimacy and enters the system with a legal account. At the same time, the system obtains the authorization to log in to the account and takes appropriate actions in accordance with the authorization level. Once the account and password are inaccurate, the system will refuse to log in to the account.

3.2 Model Management Module

- (1) Classification management
Model categories include hard fittings, kitchens and bathrooms, lamps, furniture, appliances, soft fittings and tools. When the user logs in to the system, the model classification appears in the model library menu in the form of a list, and the user can quickly find the desired model through the model classification.
- (2) Model query
After logging in, users can search for models by model name or model type and view the query results.
- (3) Model attribute management
Model attributes are divided into three categories: geometric attributes, relative change attributes and product attributes. The geometric attributes mainly include the size of the model, and the related change attributes mainly include the displacement, angle, and zoom ratio of the model. Product attributes are divided into model name, brand and specifications. The model attribute management is mainly the management and display of model attributes. After logging in to the system, the user selects a model from the current scene, and displays the model properties of the module in the menu.

3.3 Design Plan Management Module

- (1) New design plan
After the user logs in to the system, before starting any design work, open an existing drawing plan or create a new drawing plan. Select Open Design in the design management menu to display the opened design page and list the design drawings belonging to the current user.
- (2) Save the design plan
After finishing the design work, click the "Save" button to save the current design plan. When saving a drawing plan, the name of the drawing plan defaults to the name of the drawing plan house type.
- (3) Delete the design plan
Display a list of all shapes drawn by the current user. The user can open or delete the drawing plan in the list, and click the delete button behind the specific drawing plan to delete the drawing plan.

4 Discussion

4.1 System Performance Test

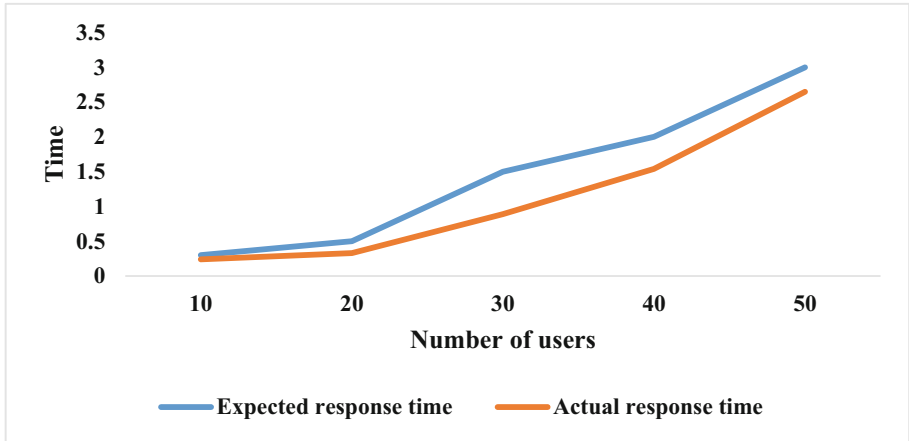


Fig. 1. Performance test results

It can be seen from Fig. 1 that when the number of concurrent users reaches 50, the response time of the system designed in this article is 2.65 s, and the actual response time is within the expected response time, which meets the needs of this article.

4.2 User Satisfaction Survey

In this article, an interior design company selected customers to conduct related surveys. After entering the page, the customers tried out the different functions of the system, focusing on the appearance, ease of use, practicality and interactivity of the system interface. Random questionnaire surveys were conducted, and the survey results were as follows Table 1 shows.

Table 1. Questionnaire analysis form

	Excellent	Good	Generally	Poor
Interface display effect	15	54	25	3
Ease of operation	12	36	22	2
Convenience	38	43	22	0
Interactivity	17	56	15	5
Practicality	14	70	16	7

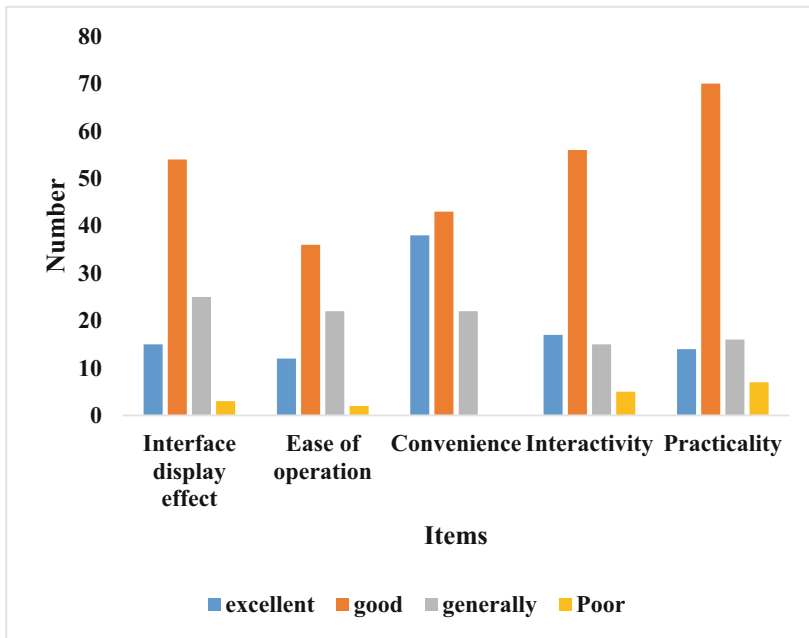


Fig. 2. Questionnaire analysis form

It can be seen from Fig. 2 that the user's evaluation of the system in this article is still very good. Most users find it interactive and are satisfied with the page display effect. The convenience and practicality of the display system can also make users feel strong. Interest, generally speaking, has good practicality.

5 Conclusions

With the continuous development of science and technology, the rapid development of computer technology has provided abundant opportunities for the scientific and modernization of the interior design industry. To fully analyze and play these possibilities in the field of interior design is also an inevitable requirement of the times for interior design.

References

1. Tokumasu, F., Hayakawa, E.H., Fukumoto, J., et al.: Creative interior design by Plasmodium falciparum: lipid metabolism and the parasite's secret chamber. *Parasitol. Int.* **83**(4178), 102369 (2021)
2. Yang, L.: Practical analysis on the integration of fair-faced concrete decorative elements into modern interior design. *J. Landscape Res.* **12**(02), 9–12 (2020)
3. Koger, M.: Updated interior design services contract documents coming soon. *Eng. News Rec.* **282**(9), 50 (2019)

4. Kazuya, Y.: Computer-aided design/computer-assisted manufacture derived needle guide for injection of botulinum toxin into the lateral pterygoid muscle in patients with oromandibular dystonia. *J. Oral Facial Pain Headache* **32**(2), e13–e21 (2018)
5. Daemei, A.B., Safari, H.: Factors affecting creativity in the architectural education process based on computer-aided design. *Front. Archit. Res.* **7**(1), 100–106 (2018)
6. Khan, R.U., Zhu, W.: Computer-aided design and property prediction of novel insensitive high-energy heterocycle-substituted derivatives of cage NNAHP. *J. Mol. Model.* **26**(9), 1–12 (2020). <https://doi.org/10.1007/s00894-020-04513-2>
7. Wen, L., Vandewalle, K.S., Walker, S., et al.: A survey of US air force general dentists regarding computer-aided design/computer-aided manufacturing usage. *J. Contemp. Dent. Pract.* **21**(3), 249–252 (2020)
8. Blas, M.J., Gonnet, S.: Computer-aided design for building multipurpose routing processes in discrete event simulation models. *Eng. Sci. Technol. Int. J.* **24**(1), 22–34 (2021)
9. Li, K., Li, J., Du, W., et al.: Computer-aided design and manufacturing cutting and drilling guides with prebent titanium plates improve surgical accuracy of skeletal class III malocclusion. *Plast. Reconstr. Surg.* **145**(5), 963e–974e (2020)
10. Buonamici, F., Furferi, R., Governi, L., et al.: A practical methodology for computer-aided design of custom 3D printable casts for wrist fractures. *Vis. Comput.* **36**(2), 375–390 (2020)
11. Zhong, Z., Peng, Z.M., Hao, D.M., et al.: A comparison between X-ray imaging and an innovative computer-aided design method based on weightbearing CT scan images for assessing Hallux Valgus - ScienceDirect. *J. Foot Ankle Surg.* **60**(1), 6–10 (2021)



Design and Research of Public Rental Housing Rent Pricing System Based on Mobile Internet

Qiong Zhang¹ and Hui-yong Guo²(✉)

¹ School of Economics, Wuhan Donghu University, Wuhan, Hubei, China

² School of Management, Wuhan Donghu University, Wuhan, Hubei, China

Abstract. The 21st century is a convenient, environmentally friendly. China's economy has entered a stage of rapid development, and the process of urbanization has greatly accelerated. China's real estate industry has also entered a period of rapid growth, and has become an important engine to promote economic development, and has become more and more closely related to the overall development of the national economy and the lives of ordinary people. With the continuous development of the real estate industry followed by rapid development, the real estate integrated service market is constantly maturing, resulting in commercial housing sales, leasing markets, stock housing sales, and real estate mortgage and development appraisal, intermediary, property management and other transaction markets. Based on this, this paper proposes a public rental housing rental pricing system based on MI. This article adopts the B/S architecture, uses the main technologies such as J2EE, JSP, SQL, and MySQL database to design and implement a public rental housing rental pricing system based on the MI. Experiments show that the system designed and implemented in this paper has passed the software test and can run well on the server side, and there is no stalling phenomenon when the system.

Keywords: Mobile internet · Public leasing · Housing rent · Pricing system

1 Introduction

At this stage, our country is moving towards a well-off society as a whole, the people's food and clothing problem has been basically solved, the economy has continued to develop in a healthy manner, and the real estate market that has followed is also rapidly rising [1, 2]. According to relevant data, as of the end of 2010, the living space of urban and rural people in my country reached 31 square meters. Even so, due to the large gap between the rich and the poor, there are still groups without houses and snail-living groups in some areas. In the future, the real estate lease transaction market will further develop steadily [3, 4]. With this strong market demand as the background, China's current 88,000 large, medium, small, and leasing companies will develop coordinately. It is necessary to further improve housing through the use of information system management, big data and blockchain applications. The scale, intensification and professionalism of leasing companies [5, 6].

From the perspective of the entire world, the housing security system is affected by the level of political and economic development, but each country has its own particularities. Here are a few representative and distinctive researches on national housing security systems: The most distinctive housing security system in the United States is to maintain the effective implementation of the housing security system through laws [7]. In the 1930s, a series of housing security systems were promulgated and implemented in the United States through legislation. Through a high level of informatization, the housing of low-income groups was legally guaranteed. Due to its high level of informatization management, it was able to fully protect the housing of low-income groups. Realize housing information sharing and effectively protect the interests of low-income groups [8]. There is still a big gap in the research of public rental housing security system in our country. The research mainly focuses on urban real estate management system. As a huge housing information system, the urban real estate management information system is very complicated to use and is not suitable for the current low-income groups. Therefore, the special two-rent housing security system should be independently studied to ensure that everyone can use it conveniently and quickly [9, 10].

This system adopts the B/S architecture to realize the main functions of the comprehensive information management of public rental housing.

2 Design of Public Rental Housing Rent Pricing System Based on MI

2.1 Features of MI

Mobile Internet (MI) technology is an emerging business that uses smart mobile terminal APP and wireless communication to obtain related services. The proposed “mobile” refers to the Android smart mobile terminal, and the “interconnection” specifically refers to the interconnection of the on-site hardware equipment collection system and the web data management platform, and the web data management platform and the Android smart mobile terminal. Through layer-by-layer interconnection, a system in which Android smart mobile terminals and hardware devices are interactively connected is finally realized [11, 12].

The MI system architecture mainly includes three levels: equipment, server, and software. The equipment side contains a collection unit and a signal processing unit, which are mainly used to collect and transmit real-time data from field hardware devices. The acquisition unit mainly includes various sensor circuits, two-dimensional code recognition and RFID tags, while the signal processing unit mainly includes signal processing and conversion, as well as signal calibration and transmission processes; the server side mainly receives information from the device side and stores and updates, it mainly uses My SQL database to realize data reception and effective storage. Through a series of decoding operations, it responds to Internet customer requests in real time; the software terminal can access the database in real time to query the corresponding reliable data information, which mainly includes some applications on electronic equipment APP, such as the most commonly used APP on mobile phones, and it includes different operating systems, such as Android smart mobile terminals and IOS smart mobile terminals.

2.2 Analysis of Functional Requirements

Function analysis is mainly to define the main functions that the system needs to complete. Without functional analysis, it is impossible to know what functions the system needs to accomplish and what kind of people use this system.

Through literature and market research, it is determined that the system has the following roles: administrator, user, and homeowner. The basic functions of the three roles are explained in detail below.

- (1) Administrator: The administrator mainly has the following functional modules:
- 1) User management: manage the information of registered users and homeowners, including querying and deleting user information.
 - 2) City management: Including adding, modifying, and deleting city information. Adding a city is for future business expansion to add new city listings, so when designing the system, leave the function of adding a city alone.
 - 3) Housing information management: firstly, review the housing information released by the homeowner; secondly, view, modify, delete and other operations on the housing information; finally, when there is more housing information, it can be classified (classification includes city, housing type, homeowner, etc.) or fuzzy query housing information.
 - 4) House leasing management: Manage the houses that have been rented, including classified query, deletion, etc.
 - 5) House evaluation management: Manage users' evaluation information on rented houses, and delete malicious evaluations or irrigation evaluations. The homeowner can reply by leaving a message to ask the administrator to review the evaluation and delete the malicious evaluation. The remaining user roles can only browse the evaluation information of the relevant house.
 - 6) Announcement management: including modifying and deleting announcements, adding announcements to the sub-function modules below it, and because of cooperation with other websites, you can publish some information on other housing rental websites.
 - 7) Message management: mainly includes replying to users and homeowners' messages, for everyone to communicate and use. Secondly, delete illegal messages and so on.

Through the above analysis, the administrator includes the following use cases: user management, city management, housing information management, housing lease management, announcement management, and message management. Among them, the housing evaluation management use case relies on housing information management and housing lease management.

- (2) User:
- 1) View all the house information, the house information displayed on the homepage has been approved by the administrator.

- 2) View the information of leased houses. Including the type of rental housing, publisher, tenant, rent and other information, as well as comments on housing information.
- 3) User message: mainly for online communication.
- 4) View the announcement. The user's use case diagram is shown in Fig. 1.

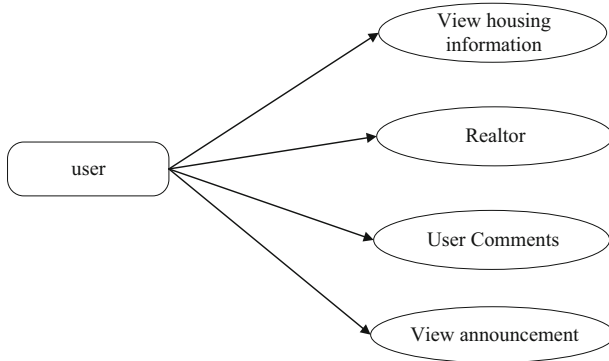


Fig. 1. User use case diagram

2.3 Related Algorithms

(1) Variance

The formula for calculating the degree of deviation between the signal strength and the overall average value of the variance is as follows:

$$VAR = \frac{1}{N-1} \sum_{i=1}^N (x(i) - \bar{x})^2 \tag{1}$$

(2) Absolute Integral Average

The formula for the average intensity of the ultrasonic audio signal is as follows:

$$AVA = \frac{1}{N} \sum_{i=1}^N |x(i)| \tag{2}$$

3 System Test

3.1 Test Environment

After coding and debugging the entire system, the designed and developed system needs to be deployed and tested. The code of the house rental management system based on Internet+ is deployed on Tomcat7 integrated with MyEclipse. The hardware configuration is as follows: one ACER5742 computer, Intel i5 processor, and Windows7 operating system. Use Google browser to access the system.

3.2 System Function Test

Functional testing, as the name implies, is a process of pre-trialing the specific functions developed by the software before it is put into application, and finally achieving user satisfaction. The following is the specific operation content and process:

- (1) Menu inspection: As the name implies, it is whether the commands and execution of each program of software development are consistent, and whether the commands can be operated normally.
- (2) Associated content inspection: Software development is not a single program to perform operations, so in the testing process, we should also pay attention to the inspection and testing of other associated programs.
- (3) Button test: A series of execution functions of buttons in the design process must be tested.

4 System Test Analysis

4.1 Stress Test

Initially, the system performed 3000, 5000, and 7000 state-of-the-art tests to ensure the printing performance of different user devices. And in the actual application process, an MI-based rental system was selected as the key function with the highest use frequency and planning requirements: completion of homework test scores, etc., and return of test case information.

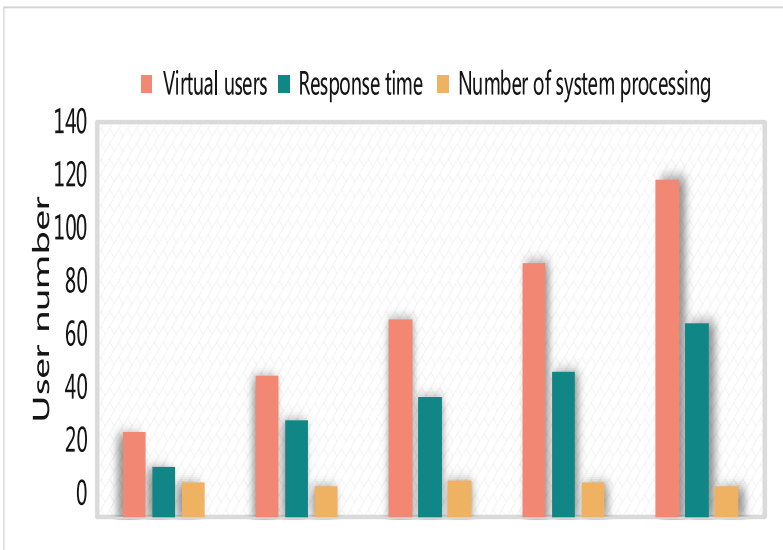


Fig. 2. Stress test statistics

As shown in Fig. 2, through the analysis of the stress test scale, when the number of visits to the system increases or decreases, the system's simultaneous response increases, but even with 7000 visits per minute, the response to the system is also a clock. This program is not bad. The actual response time does not exceed 3 s, which can fully meet the requirements. On the other hand, it can be determined that the program fully meets the user login requirements. Check the number of times the system processes per second, the actual operating mode will not change due to the increase in CPU usage, and the performance will be better.

4.2 Administrator Test

The administrator test is shown in Table 1.

Table 1. Administrator test

Test function	Testing process	Test results
Login function	Enter the URL, click to enter the administrator login page, enter the account: admin and password 111111, and click OK	Administrator successfully logged in
User management function	After logging in successfully, click User Management to view the user's background information and modify or delete a user	User management function is normal
Housing information management function	Click House Management, add a house and modify its information, and finally delete it	Housing information management function is normal
Housing evaluation management function	Click Housing Evaluation Management to view the evaluated housing and the rental object of the message	Housing evaluation management function is normal

The three roles of administrator, homeowner, and user are tested in the Google browser. All functions have passed the test, and the business interaction between different roles is normal. When a message is not deleted, the business logic of other roles is in disorder. The various functions between users can interact well. If one party changes, the other party's business logic will also change accordingly. In general, the system designed and implemented in this paper has passed the software test and can run well on the server side, and there is no stalling phenomenon when the system operates the business process.

5 Conclusions

With the rapid economic growth and urbanization, more and more people are pouring into cities, and soaring housing prices have discouraged many low- and middle-income

earners. Therefore, people need to rent houses as collateral, which will promote our rental market. At the same time, product promotion is a prerequisite for increased competition among leasing companies. At present, the company can only use related technologies to reduce user rental time, save rental procedures, and reduce costs. Creating systems and websites that are simple to operate and easily complete leasing through the Internet can attract a large number of small and medium-sized users, thereby increasing the brand awareness and initiative of leasing companies. Therefore, renting a house is a big deal. At the same time, “Internet+ ” is developing rapidly. Combining this technology with the traditional leasing management system also brings certain benefits to the application of “Internet+ ” in the traditional housing leasing system.

Acknowledgments. This research is funded by the (Wuhan Donghu University, Wuhan, China) Youth Foundation Project in 2020 (Sustainable Recycling Mechanism of Wuhan’s Public Rental Housing Construction Funds) (Item Number: 2020dhsk006) Project Leader: Qiong Zhang.

References

1. Wulandari, D.W., et al.: Social living environment in Kost private rental housing based on tenants’ experience. *J. Archit. Plann. (Transactions of AIJ)* **84**(757), 527–536 (2019)
2. Kim, K., Ran, K., Ko, K.: Measuring the effect of public rental housing residency using matching methods: a focus on a comparison of the mental health of public housing tenants and private housing tenants. *Korean J. Public Adm.* **57**(4), 255–282 (2019)
3. Li, H., Zhang, S., Shi, J., et al.: Research and design of intelligent learning system based on recommendation technology. *Control. Intell. Syst.* **47**(1), 43–49 (2019)
4. Sobri, M.J., Ijab, M.T., Nayan, N.M., et al.: Design of mobile heart disease application based on users’ perspective using design science research method. *J. Inf. Syst. Technol. Manage.* **5**(18), 61–75 (2020)
5. Park, J.Y., Jeong, J.J., Park, W.J.: A study on the community space that affect the public rental housing satisfaction determinants. *KIEAE J.* **16**(1), 95–101 (2016)
6. Gan, X., Zuo, J., Baker, E., Chang, R., Wen, T.: Exploring the determinants of residential satisfaction in public rental housing in China: a case study of Chongqing. *J. Housing Built Environ.* **34**(3), 869–895 (2019). <https://doi.org/10.1007/s10901-019-09691-x>
7. Persson, L., Persson, A., Sam, C.: Implementation of the strategic approach to international chemicals management in Cambodia: effects of regime design. *Int. Environ. Agreements Polit. Law Econ.* **16**(1), 1–20 (2016)
8. Wang, Y.H., Correia, G., et al.: Understanding travellers’ preferences for different types of trip destination based on MI usage data. *Transp. Res. Part C Emerg. Technol.* **90**(May), 247–259 (2018)
9. Kim, H., Jang, S.: Analyzing the location of happy-rental housing based on university students’ commuting trip. *J. Korea Plann. Assoc.* **52**(5), 89–112 (2017)
10. Wijburg, G., Aalbers, M.B., Heeg, S.: The financialisation of rental housing 2.0: releasing housing into the privatised mainstream of capital accumulation. *Antipode* **50**(4), 1098–1119 (2018)
11. Kate, C.: Rental housing at a ‘Turning Point.’ *Prof. Builder* **83**(2), 11 (2018)
12. Moon, O.: Living on your own: single women, rental housing, and post-revolutionary affect in contemporary South Korea. *Am. Anthropol.* **118**(2), 454–455 (2016)



Design of Interior Design Database Based on Ant Colony Algorithm

Yuhan Zhang^(✉)

School of Arts and Humanities, Yunnan Technology and Business University,
Kunming, Yunnan Province, China

Abstract. Interior design can be said to be the personal pursuit of contemporary people after the improvement of living standards. People's requirements for information technology processing are getting higher and higher. In the interior design work, the design database plays an important role in it. The purpose of this article based on ant colony algorithm research and design of interior design database is to organize interior design resources and provide people with some interior design methods and opinions, in order to achieve beautiful housing and meet users' expectations. This article mainly applies the experimental method and the comparative method to compare and analyze the unit test and integration test results, and get the corresponding data results. The data shows that in the database system designed in this paper, its performance indicators can basically reach 80%, and its security performance has a high level, which can protect user privacy.

Keywords: Ant colony algorithm · Interior design · Database research · Design method

1 Introduction

In the modern design industry, the importance of design is self-evident. A good design plan can often bring good economic benefits to the designer. The method of interior design is no longer purely manual drawing, but uses modern information technology to update the interior design method. The ant colony algorithm has a great effect in the database, so the research on the interior design database can be carried out from the ant colony algorithm.

There are many theoretical results in the research and design of interior design database based on ant colony algorithm. For example, some people have proposed that the interior design of a building is to decorate and arrange the space according to their own preferences to enhance the overall connotation of the building [1, 2]. Some people think that an intelligent data processing system for interior design needs to understand the field of data processing technology, including the concept of connecting interior design [3, 4]. In addition, in order to break through the limitations of traditional interior design education and cultivate high-tech talents, based on the world's scientific and technological network information, some people develop and build a library of interior design disciplines [5, 6]. Therefore, this article based on the ant colony algorithm has

certain reference value and reference significance for the research of interior design database.

This article first studies the interior design database and elaborates on it. Secondly, the corresponding research and understanding of the data warehouse are carried out. Then it analyzes the application of ant colony algorithm in interior design database. Finally, perform software testing on the system and obtain data results.

2 Research and Design of Interior Design Database Based on Ant Colony Algorithm

2.1 Interior Design Database

The database is used to store data, it can be said to be a kind of “soft” system in the computer. The interior design database is a complex process, which involves many aspects, including the physical layer, application software layer and so on. It needs to be researched and analyzed in actual operation. In theory, it is to decompose the function of the computer system into several small unit modules to form the entire system; from a practical point of view, it is an engineering information system (MIS) model established according to certain rules [7, 8].

The interior design database is used by designers to classify, sort and process the problems to be solved in the system according to certain requirements. It mainly includes the realization of specific functions according to needs, and users can share some data information. The interior design database is mainly used to store the information of the design works produced by interior designers. Due to the huge amount of data in these programs, it needs to be carried out and then shared on the computer. The interior design database can not only store a large amount of content about user behavior rules, needs, etc., but also analyze, sort and count these data, and provide designers with a certain basis for decision-making. The purpose of researching the interior design database is to classify its design content through the database, and then assign it to each module. Through these records, we can conveniently perform various queries, statistical analysis and other tasks. At the same time, it can classify user needs and provide corresponding solutions [9, 10].

In our design, in order to better realize the design tasks, we need a data management system with clear levels and clear division of labor. The most important of these is to record and query the underlying information. This can effectively avoid unnecessary losses due to the inefficiency or waste of resources caused by the independence of various departments. Middle level data: This is an important level concept. It mainly includes business logic, operation process and management interface. At the same time, it also contains some comprehensive factors such as knowledge and experience accumulation in software development technology and management methods. The other level is used to provide the specific implementation of the corresponding service node and the database connection relationship in the corresponding subroutine and the related information storage layer, access management layer, and query statement management control of the attributes between them [11, 12].

In order to ensure the security and reliability of the database, the data needs to be encrypted, so as to ensure the security of the system. The first is the establishment of the underlying information table. The second is the interaction between the user side and the upper node, and the server side. Finally, a hierarchical structure model is constructed between the server side and the application layer to realize data sharing and exchange functions.

The design of this article is a database based on ant colony algorithm. Its purpose is to optimize the performance of the entire model. This process is to analyze and research calculations according to each underlying application software level. Then through the integration and summary of the information content of the bottom database tables, subroutines, and all data storage layers on the upper application programs in each module, the corresponding conclusions can be drawn.

2.2 Data Warehouse

With the development of information and decision support systems, data warehouses have emerged. As an auxiliary tool of decision support system, data warehouse system mainly includes back-end data warehouse service and front-end analysis service. The data warehouse is generated by the operational database. It is an analysis and application-oriented environment. It can provide real-time data and historical data for governments, enterprises and organizations, and provide support for decision-making. Its specific structure is shown in Fig. 1:

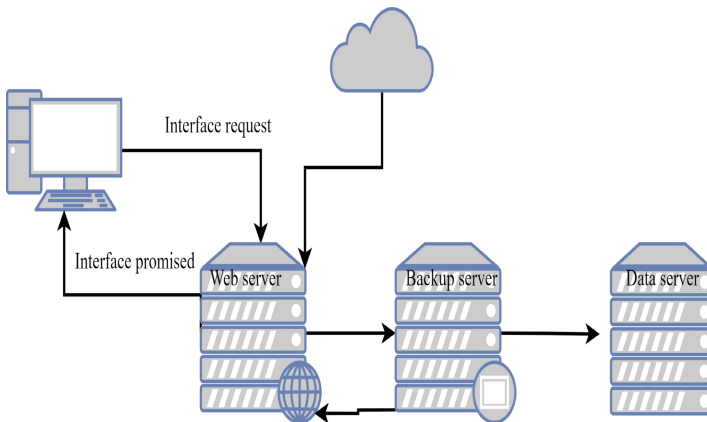


Fig. 1. Data warehouse system structure

For the data warehouse of interior design, it is undoubtedly a system oriented to meet people's needs. The interior design database is transformed into a data model by designers after analyzing system functions according to user needs. The interior design database is based on the needs of the system. According to the data and needs provided by the users, the various modules are organically combined. It mainly realizes data query and update, statistics and modification functions by collecting and analyzing information.

The database technology mainly has the following several:

Data storage technology. Compared with traditional data warehouse storage, users can freely design the type of data storage. Therefore, when designing the system data storage organization, in addition to considering how the storage accesses the data, the impact on storage performance and the impact of intermediate caching on storage access should also be considered. **Data access technology.** According to different strategies, there are many ways to implement indexes. At this stage, the request needs to be repeated multiple times to complete the request. The technology to process the request. The query optimization algorithm is also mainly to reduce the number of hard disks and increase the speed. **Backup and recovery technology.** To avoid the loss of memory data due to a crash, a backup should be kept on non-volatile memory.

2.3 Application of Ant Colony Algorithm in Interior Design Database

Ant colony algorithm can be regarded as a distributed self-organizing multi-agent system with good global search capabilities. With the continuous accumulation of pheromone, more and more pheromone hinders better solutions, and ants are increasingly looking for solutions that are close to the best solutions. It can be seen that the ant colony algorithm embodies a mechanism different from the conventional algorithm and has long-term research value from the point of view of system science. In the optimization process, the ants use the residual information on the path and the heuristic information on the path to calculate the probability of state transition. The formula is as follows:

$$P_{mn}^l = \begin{cases} l \notin allowed_l, & \frac{(\mu_{mn})^\sigma (\lambda_{mn})^\nu}{\sum_{t \in allowed_l} (\mu_{mn})^\sigma (\lambda_{mn})^\nu} \\ l \in allowed_l, & 0 \end{cases} \tag{1}$$

Among them, $allowed_l = \{d = tabu_l\}$ represents the nodes that ant l is allowed to select in the next step.

In order to avoid overwriting heuristic information with too much residual information, its update rule is as shown in formula (2):

$$\zeta'_{mn} \leftarrow (1 - \psi)\zeta_{mn} + \Delta\zeta_{mn} \tag{2}$$

Among them, ψ represents the information volatilization factor. When the ant traverses all the elements and completes a cycle, it updates the pheromone according to the global information. The model has a good effect in solving the data problem.

In order to make the various modules of the interior design database play a better role, the data must be classified reasonably and effectively in order to improve the efficiency and quality of the entire system. The central database is used to store the required information in the data warehouse, and can realize the management of various physical resources. There are three main parts: basic information, files and applications. The carriers are readable formats such as CDs and disks; files include various types of tables and various documents, graphic documents and other formats. Tools are used to access data; application software can obtain related software or Resource information on the

hardware device. The indoor database of this design is based on “system” and “module”. The interior design database module based on ant colony algorithm mainly includes: data management, query function and user authority. Because the ant colony algorithm has better fault tolerance, robustness and fast characteristics. So this article uses this algorithm to solve the problem of interior designers’ high sensitivity to data parameters. The interior design database based on ant colony algorithm is mainly composed of two parts. One is to design from the system requirements; the other is to write functional modules and programs generated during the entire software development process.

3 Software Testing

3.1 Test Purpose

Software testing is to be able to find some problems in the software program, to effectively and accurately solve the problem after finding the problem, and to find out whether the various assembly levels of the software can meet the requirements. Testing the software function template is the main way.

3.2 Test

Unit testing is to discover possible defects and errors within each module. Mainly for the smallest part of the program module of the software design, start from the internal functional structure to carry out a certain inspection and test work. The main method of this unit test is to write unit tests while developing the system at a stage of system development, and constantly test the units to verify the software.

In addition to unit testing, integration testing is usually an assembly test. On this basis, all modules should be assembled into a system in accordance with the design requirements. In terms of testing, we need to consider some more important issues. The first thing to consider is the data of the test module interface to avoid data loss. The second consideration is whether the function of the module can work properly with the function of another module without causing unsmooth problems. Whether the combination of each sub-function can meet the requirements of the parent function. Once again, we must consider the size of the error of the module, whether all the accumulation of errors will be magnified to affect the possibility of acceptance. Finally, consider whether there is a problem with the global data structure.

3.3 Test Plan

The login module includes logging in to the system, logging out of the system and changing user passwords. A legitimate user enters the system, and the system receives the login account information so that he can perform the authorization of the appropriate operation. If the password is incorrect, the system will refuse to log in to the account.

4 Test Results and Test Evaluation

4.1 System Performance Analysis

According to the test experiment, this system can meet most of the conditions required in the system analysis, and can also meet the requirements of the functional design in the system design, which means that it has achieved the purpose we anticipated and envisaged. The specific situation is shown in Table 1:

Table 1. System performance analysis

	Unit testing	Integration testing
Compatibility	0.79	0.87
Safety	0.89	0.93
Easy to use	0.81	0.84
Compressive resistance	0.85	0.91

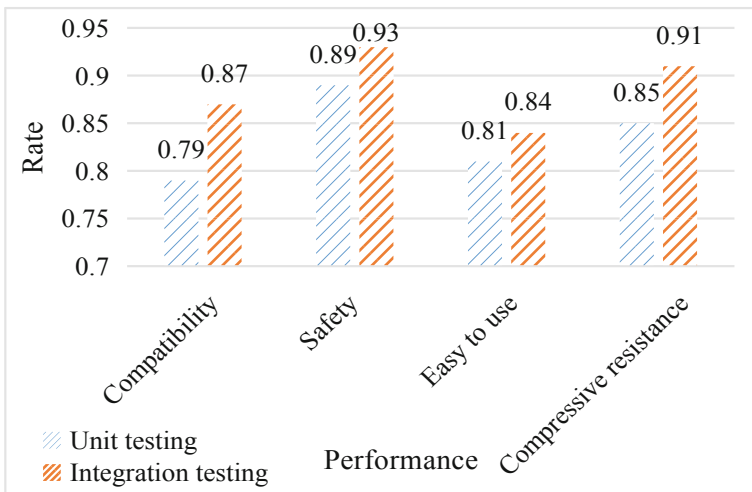


Fig. 2. System performance analysis

As shown in Fig. 2, we can see that in this system, its safety performance is the highest, followed by its pressure resistance. The third place is compatibility, and the last place is ease of use. However, the results of unit testing and integration testing are different. The general data of unit testing is lower than that of integration testing.

5 Conclusion

In this article, the interior design database model is composed of a data warehouse and multiple data stores. When designing a database, it is not only necessary to consider

the software functional requirements, performance indicators and other issues. It is also necessary to make full use of modern communication technology to realize the needs of information sharing and exchange among users. In this process, new challenges were raised to software technology and hardware facilities. In this paper, the use of ant colony algorithm for database research is conducive to making full use of modern information technology to process data. Through experiments, it is found that the system designed in this paper has good performance and reliability. But it needs to be improved in terms of its ease of use.

References

1. Chen, Y.-Q., Guo, J.-L., Yang, H., Wang, Z.-Q., Liu, H.-L.: Research on navigation of bidirectional A* algorithm based on ant colony algorithm. *J. Supercomput.* **77**(2), 1958–1975 (2020). <https://doi.org/10.1007/s11227-020-03303-0>
2. Yin, Z., Du, C., Liu, J., et al.: Research on auto-disturbance-rejection control of induction motors based on ant colony optimization Algorithm. *IEEE Trans. Ind. Electron.* **PP**(4), 1-1 (2017)
3. Cheng, C., Zhou, X.D., Gao, M., et al.: Research on monopulse forward-looking high-resolution imaging algorithm based on adaptive iteration. *Defence Technol.* **16**(1), 158–171 (2020)
4. Mun, J.-E., Kim, J.-H., et al.: The study for securing reproducibility of experimental method in papers of spatial image evaluation-focusing on the papers of domestic journals. *J. Korean Inst. Inter. Des.* **26**(4), 93–102 (2017)
5. Liang, D.: Design and implementation of teaching resource library for interior design based on cloud platform. *Electron. Des. Eng.* **025**(019), 75–78 (2017)
6. Schwenzfeier, T.: Interior-design Mit Internationalen Top-Playern. *Textilveredlung textilchemie* **5**(11–12), 6–8 (2017)
7. Radley, L.: Interior design. *Commer. Mot.* **227**(5740), 32–37 (2017)
8. Zheng, L., Wang, Z., Zhao, Z., et al.: Research of bearing fault diagnosis method based on multi-layer extreme learning machine optimized by novel ant lion algorithm. *IEEE Access* **PP**(99), 1-1 (2019)
9. Wang, R., He, G., Wu, X., et al.: Multicast optimization and node fast location search method for wireless sensor networks based on ant colony algorithm. *J. Digit. Inf. Manage.* **15**(6), 303–311 (2017)
10. Bamdad, K., Cholette, M.E., Guan, L., et al.: Ant colony algorithm for building energy optimisation problems and comparison with benchmark algorithms. *Energy Build.* **154**(nov.), 404–414 (2017)
11. Prakash, N., Rajalakshmi, M., Nedunchezian, R.: Optimized energy aware routing based on suitable based antlion group with advanced algorithm (SA-AOA) in wireless sensor network. *Wireless Pers. Commun.* **113**(1), 59–77 (2020). <https://doi.org/10.1007/s11227-020-07178-5>
12. Li, S., Cai, S., Li, L., Sun, R., Yuan, G.: CAAS: a novel collective action-based ant system algorithm for solving TSP problem. *Soft Comput.* **24**(12), 9257–9278 (2019). <https://doi.org/10.1007/s00500-019-04452-y>

CPS Systems, System Integration, and Real-World Applications



Application of Computer BIM Software Technology in Building Information Model

Xiu Luo¹(✉) and Amar Jain²

¹ College of Materials and Environmental Engineering, Chengdu Technological University, Chengdu 611730, China

² Madhyanchal Professional University, Bhopal, India

Abstract. Using traditional information communication and management methods is very easy to make project cost management out of control. The BIM (Building Information Model) building information model can integrate the information of all stages of the project, promote the participation and cooperation of the project main body, and carry out cost management before the actual construction of the project, so as to realize the cost management of the whole process from design to construction and operation, effectively. It avoids the situation that the budget will be exceeded after the project is completed. The purpose of this article is to study the application of BIM software technology in engineering cost specialty. In response to the problems in cost management in our country, many scholars have proposed to introduce building information models into construction projects. In price management, the distributed BIM model can be used to solve the problems of BIM software technology. To establish a cost model for a specific project, the designer converts the traditional two-dimensional engineering drawing into a three-dimensional model, and manages the project cost in detail to realize the true cost management of the entire process. This article uses the questionnaire survey method to explain the application research of BIM software technology in the engineering cost specialty. The experimental results show that in 2012, 2015, 2017 and 2019, the cost of the project using BIM technology was reduced by 4%, 6%, 10%, and 17%, respectively, and the calculation efficiency of the project increased by 7%, 10%, 12%, 15%. The research results show that BIM software technology has a practical and reliable effect on engineering projects in the construction industry, which can save project costs and improve the calculation efficiency of engineering projects.

Keywords: BIM software technology · Project costs · Applied research · Construction industry · Cost management

1 Introduction

In some developed countries in Europe or America, their construction industry informatization has developed very well, and the frequency of application of information technology in the construction industry is getting higher and higher, which is mainly reflected in the extensive use of building information models. The use of building information

models can not only improve the calculation efficiency of engineering projects, but also effectively control costs, which has been widely concerned and recognized in the industry. With the support of BIM technology, architectural design products can transition from two-dimensional to three-dimensional or n-dimensional, and people can observe the design results based on BIM software technology. Through construction simulation, designers can find out the unreasonable factors in the design earlier, thereby greatly reducing the rework problems caused by engineering changes and the cost increase caused by rework. Owners, designers, supervisors, and contractors use building information models to make communication between them easier [1]. In the entire project design process, the design plan can be continuously optimized, and cost management can be carried out before the actual construction of the project, so as not to exceed the actual cost after the project is completed.

As an important production tool, BIM software technology has always been a hot research topic in the construction information industry. Many scholars have studied this. M Woof once mentioned this technology. A software that can improve the efficiency of construction projects is made based on this technology. The software provides online access to multiple layers. With real-time project maps and geographic data functions, it provides a web-based viewer that enables the construction team to access the latest maps, thus saving valuable time and money [2]. Xiongjue has tried to use BIM technology to model the Sun Valley, including architectural modeling and structural analysis. Through the analysis of the structure of the Sun Valley, it is proved that writing script codes in Rhino software can solve the problems in the modeling, and the stability of the model is analyzed. This method is combined with Rhino, Revit, Midas and other software to solve the modeling and calculation problems of complex surface structures [3].

Understand the related concepts of BIM technology, including the basic principles and characteristics of BIM technology. Through the application of BIM technology, this article has changed the Chinese project cost management model, thus discovering its advantages and promoting the development of China's project cost management model; then, according to my country's existing cost management model, find out the BIM software technology in it Played a role. This paper uses a questionnaire to reflect the impact of BIM software technology on the project cost. The questionnaire is mainly for experimental research from two aspects, namely the impact of BIM on the project cost and the impact of BIM on the calculation efficiency of the project. Finally, this article summarizes the research, puts forward the shortcomings of this article, and prospects the application prospects of BIM software technology in my country's engineering cost specialty.

2 Research on BIM and Engineering Cost Management Theory

2.1 Basic Principles of BIM Technology

The key to BIM is information. The BIM model is basically composed of a single component or object. Use parameters to represent the geometric information, physical characteristics, structural requirements and cost information of all components or objects, and then use spatial topological relations and 3D Boolean operations to classify this information, and store it in a huge database, and finally form a digital model [4]. The

BIM project system can change relevant information at any time according to the changes of the project. This will not only solve the problem of information redundancy, but also all stakeholders can exchange information more conveniently. The BIM model is mainly based on digital technology and has unique attributes that can express and distinguish the key components of the model through parameters. It has object-oriented characteristics, can receive data, process data and transmit data to other objects [5]. BIM technology can use parameterized basic elements to build a model, provide the required information for other related construction work, and can also receive information feedback, and make the model more complete through information processing.

2.2 Project Cost Management

From a narrow perspective, project cost refers to the construction price of a certain project. In a broad sense, project cost includes construction project cost, installation project cost, municipal project cost, power project cost, water conservancy project cost, communication project cost, etc. We can understand project cost management from two aspects. First of all, from the perspective of the construction party, they believe that the management of project costs involves the reproduction of fixed assets, the production of corresponding intangible assets, and initial operating costs. These project costs are not only the fixed asset investment of the construction project, but also involve “sales”. It is also the basis for investors’ pricing in the market. Secondly, from the contractor’s point of view, project cost management refers to the actual or estimated price of the project and the total price of the construction project formed in the land market, technical labor market, and equipment market [6, 7]. Project cost management is similar to other trade activities. Under the control of the socialist commodity economy and market economy, the cost is usually regarded as the project contract price.

2.3 Advantage Analysis of BIM Software Technology in the Application of Engineering Cost Management

BIM software technology improves the efficiency of engineering quantity calculation: the core content of engineering project cost management is engineering quantity, which is the basis of all cost management activities, such as cost calculation, engineering bidding, business negotiation, contract signing, progress payment, etc. [8]. According to national standards and related calculation rules, BIM technology-based engineering quantity calculation software and entity deduction operations are carried out, which greatly improves the accuracy of engineering quantity calculations. At the same time, it can automatically output electronic files for information exchange, sharing, and long-distance transmission. And permanently save project participants. In addition, different professional participants of the same project do not need to re-establish the model, and only need to input professional data information in the constructed model to obtain the calculation results. The database function of the BIM model is used to store the information of each component of the project. Cost personnel can freely extract the component information related to the project during the calculation process, which not only improves the calculation efficiency, but also reduces the subjective errors of the personnel in identifying the component information the possibility of obtaining more

objective and accurate data. In addition, with the continuous improvement of cloud computing technology, we can intelligently check the number of BIMs and improve the accuracy of the model.

BIM improves the collaboration capabilities of project participants. The application of technology in project cost management realizes real-time dynamic analysis, sharing and collaboration of horizontal and vertical information. The realization of this function is of vital importance to the cost control of all parties in the project and to improve the transparency of the construction market. It also provides good technical support for realizing the full life cycle management of project costs. It improves the ability of early management and control of project cost: BIM calculation software can quickly and accurately calculate the quantity, and designers can obtain project information and data in time, thereby improving the ability of early management and control of project cost [9]. The model established by BIM can connect components and cost information data, and can change the content intuitively and concisely, and then get the result [10]. The designer can grasp the changed information in time and understand the impact of the design plan change on the cost, which also facilitates the designer to control the cost of the project in the early design stage of the project.

2.4 Cluster Analysis Algorithm

Cluster analysis is a statistical method used to analyze the classification of indicators or samples. There are mainly the following kinds of commonly used analysis.

(1) Euclidean Distance

$$d(i, j) = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (1)$$

In this equation, is the value of the k-th parameter in the i-th data sample in the data space, and is the value of the k-th parameter in the j-th data sample.

(2) Squared Euclidean Distance

$$d(i, j) = \sum_{k=1}^m (x_{ik} - x_{jk})^2 \quad (2)$$

This equation represents the sum of the squares of the differences between the two data samples and m variables. In this equation, is the value of the k-th parameter in the i-th data sample in the data space, which is the j-th parameter The value of the kth parameter in data samples.

(3) Interval Distance

$$d(i, j) = \sum_1^m |x_{ik} - x_{jk}| \quad (3)$$

This equation represents the sum of the absolute values of the differences between the values of the m variables between two data samples. In this equation, x_{ik} is the value of the k -th parameter in the i -th data sample in the data space, y_{jk} is the value of the k th parameter in the j th data sample.

3 Application Experiment of BIM Software Technology in Engineering Cost Specialty

3.1 Data Sources

By setting up a questionnaire, compare the BIM software technology in the project cost before and after it is applied to see the changes in various indicators of the project cost after the BIM technology is used. In this experiment, a total of 80 questionnaires were collected. In order to make this experiment more reliable, this study discarded the questionnaires that used BIM technology for less than a year, leaving a total of 60 valid questionnaires. This article is based on the time you use BIM software technology to distinguish. We respectively investigated the projects that used BIM software technology in the construction industry in 2005, 2008, 2012 and 2016 and the projects that did not use the technology. The required cost and engineering calculation efficiency.

3.2 Design Method

- (1) A distinction is made between engineering projects, and engineering projects at different times are distinguished to ensure the reliability of this investigation.
- (2) Compare the changes in the cost and calculation efficiency of a project before and after the application of BIM technology.

4 Application Experiment Analysis of BIM Software Technology in Engineering Cost Specialty

4.1 Distribution of Survey Groups

Among the engineering projects surveyed, 19% were from 2012, 10% were from 2015, 21% were from 2017, and 50% were from 2019. Among them, 50% of the engineering projects were developed in recent years. From this time, the reliability of this investigation is demonstrated, and such investigation reports will be more reliable. The time distribution of detailed survey items is shown in Fig. 1.

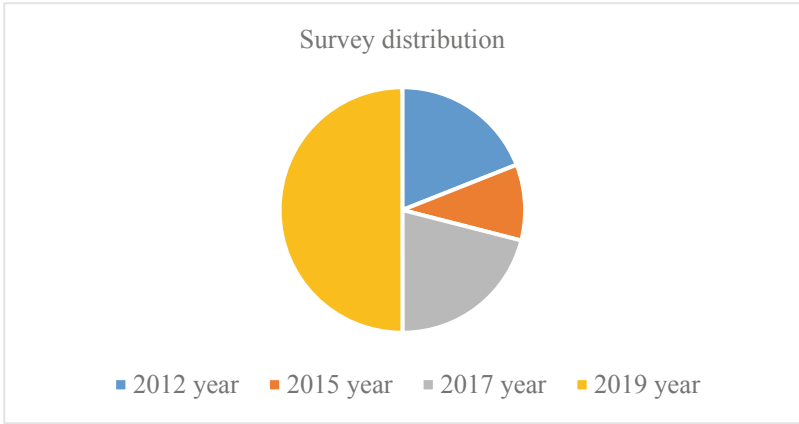


Fig. 1. Survey distribution

4.2 Changes in BIM Software Technical Indicators

The questionnaire of this study is based on two aspects, namely the cost of project cost and the problem of calculation efficiency, in order to explore an application of BIM software technology to project cost. Figure 2 shows the changes in project cost before and after the application of BIM technology.

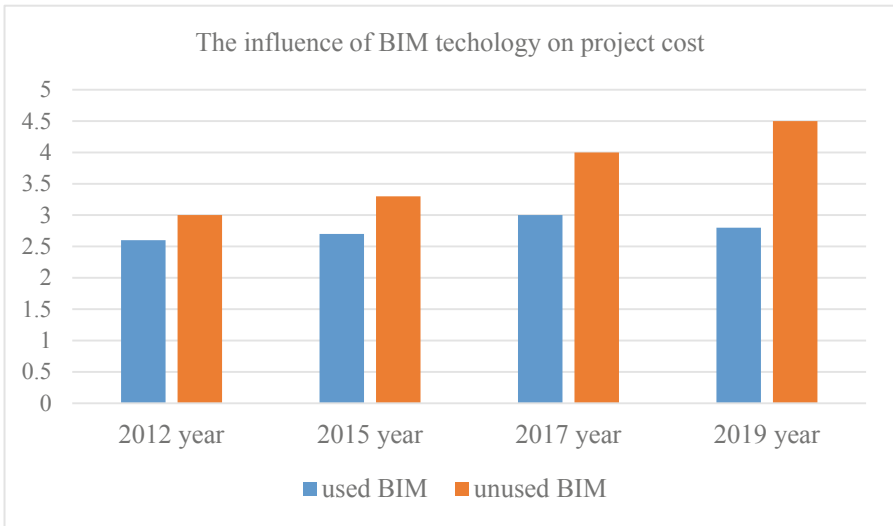


Fig. 2. The influence of BIM technology on project cost

According to Fig. 2, in 2012, the cost of engineering projects that used BIM software technology was 4% lower than that of unused ones. In 2015, the cost of engineering projects that used BIM software technology was 6% lower than that of unused ones.

In 2017, the cost of engineering projects that used BIM software technology was 10% lower than that of unused ones. In 2019, the cost of engineering projects that used BIM software technology was 17% lower than that of unused ones. From the chart, you can see that BIM technology has a greater impact on reducing the cost of the project, and this impact is gradually expanding over time. In addition, BIM technology also has a major change in the calculation efficiency of the project, as shown in Table 1.

Table 1. Influence of BIM technology on calculation efficiency of engineering quantity

Influence of BIM technology on calculation efficiency of engineering quantity	Used BIM	Unused BIM
2012 year	30%	23%
2015 year	35%	25%
2017 year	40%	28%
2019 year	45%	30%

According to Table 1, the calculation efficiency of the project using BIM software technology in 2012 was 7% higher than that of the unused project, and the calculation efficiency of the project using BIM software technology in 2015 was 10% higher than that of the unused project. In 2017, the calculation efficiency of engineering projects that used BIM software technology was 12% higher than that of unused projects. In 2019, the calculation efficiency of engineering projects that used BIM software technology was 15% higher than that of unused projects.

5 Conclusions

Although there is still a big gap between the domestic BIM technology and the developed countries, coupled with the wide variety of BIM software, the cost of installing BIM is high, but BIM software technology still has great application value in the engineering cost. BIM is an important part of project information, and the planning and actual accumulation of the project information center can enable BIM to be better applied. By cooperating with some project information centers, the company can train employees to equip them with BIM modeling, BIM software use and BIM implementation capabilities. The information center staff are enthusiastic, have a strong sense of information, and have a strong ability to use software. After one project is completed, the implementation and promotion of the second project can be carried out independently. Therefore, it is suggested that the BIM center should play a role in the implementation and promotion of BIM. When it comes to a project, it requires the strong cooperation of the project department to truly implement BIM. In this research, I have consulted a lot of related materials about BIM literature, and have a more in-depth understanding of BIM software. BIM technology is relatively powerful in some European and American countries. You

can learn more. You can allow engineering personnel to study abroad or introduce foreign experts to provide intellectual support for the development of my country's BIM and keep up with the progress of BIM in the world.

References

1. Lee, W.-L., Tsai, M.-H., Yang, C.-H., Juang, J.-R., Su, J.-Y.: V3DM+: BIM interactive collaboration system for facility management. *Vis. Eng.* **4**(1), 1–15 (2016). <https://doi.org/10.1186/s40327-016-0035-9>
2. Woof, M.: Asset management and BIM solutions. *World Highw.* **27**(3), 52–53 (2018)
3. Wang, X.: Analysis on complex structure stability under different bar angle with BIM technology. *Perspect. Sci.* **7**, 317–322 (2016)
4. Borrmann, A., König, M., Koch, C., Beetz, J. (eds.): *Building Information Modeling*. Springer, Cham (2018). <https://doi.org/10.1007/978-3-319-92862-3>
5. Huang, P., Deng, X., Huang, W., et al.: Research on safety of operation and maintenance of utility tunnel based on BIM and computer technology. *J. Phys. Conf. Ser.* **1815**(1), 012009 (16pp) (2021)
6. Daller, J., Zibert, M., Exinger, C., et al.: Implementation of BIM in the tunnel design – engineering consultant's aspect. *Geomechanik und Tunnelbau* **9**(6), 674–683 (2016)
7. Antova, G., Tanev, V.: Creation of 3D geometry in scan-to-CAD/BIM environment. *IOP Conf. Ser. Earth Environ. Sci.* **609**, 012085 (2020)
8. Munir, M., Kiviniemi, A., Jones, S.W., et al.: BIM business value for asset owners through effective asset information management. *Facilities* **38**(3/4), 181–200 (2019)
9. Kubicki, S., Guerriero, A., Lou, S., et al.: Assessment of synchronous interactive devices for BIM project coordination: prospective ergonomics approach. *Autom. Constr.* **101**, 160–178 (2019)
10. Tang, Q., Liu, Z., Chen, W., et al.: The Application of BIM in Small and Medium-sized Smart Substation—A Case Study of Prefabricated Building Design. *IOP Conf. Ser. Earth Environ. Sci.* **186**(5), 012054 (5pp) (2018)



The Design of Mental Health Information Data Management System Under the Background of Informationization

Nan Jiang¹(✉) and Rajan Singh Tiwari²

¹ Sichuan Vocational and Technical College, Suining 629000, China

² Technocrats Institute of Technology, Bhopal, India

Abstract. In recent years, due to economic development and social progress, people have faced more pressure and challenges. It is especially obvious to students who are a certain group of people. Many scientific studies in recent years have also proved that students' mental health (HM) problems are often more obvious. Therefore, it is necessary to grasp the HM of students and conduct psychological consultations in a timely manner to correct bad HM problems and students' personality problems. It is a prerequisite for teachers to educate and manage students correctly, and a prerequisite for the healthy and happy development of students' learning. This article aims to study the design of HM information data management system under the background of information technology. Based on the analysis of the problems in HM management, K-means cluster analysis and system performance requirements, the management of HM information data under the background of information technology. The system was designed, and the performance of the system was tested after the realization of each functional module. The test results show that when the number of simulated concurrent users is increased to 150, the average transaction response time is 48 ms. At this time, there is no significant delay in actual operation, which is within an acceptable range.

Keywords: Information technology · Mental health · Information data management system · College students

1 Introduction

With the development of information technology, computers have been used in HM education, consultation, evaluation and other links. Due to the unique ability of network technology, the related HM system is very suitable for applications in the field of HM, can effectively supplement the deficiencies of the traditional HM system, and is easier to be accepted by students [1, 2]. The implementation of the student HM system has made the school's HM more scientific, professional, convenient, and pertinent. It has fully grasped the HM of all students and formed a complete electronic psychological file system for campus psychological teaching [3, 4].

In recent years, information technology and artificial intelligence technology have developed rapidly. The application of expert systems, especially in the field of medical

diagnosis, has laid a theoretical and technical foundation for the development of online psychological consultation. Intelligent diagnosis technology makes full use of on-site experts, because it can effectively receive, transmit, process and use diagnostic information, simulate human experts using computers to diagnose complex problems, and can perform fast and intelligent reasoning based on the experience of various experts [5, 6]. A virtual online psychological consultation platform can protect people's privacy and facilitate the treatment of people with psychological problems. The process of psychological consultation is based on a virtual network, which can essentially avoid negative interference from reality and better solve problems [7, 8]. Some researchers have designed a system for managing students' psychological records, which can effectively manage students' HM by integrating and analyzing data from various subsystems [9, 10]. There are already many management systems related to student psychology research in our country. The content is mainly the elaboration and explanation of social psychology, but at the same time, the target users are also the entire population, and there is a lack of psychological conditions for this special student population [11, 12]. However, because the age and living environment of the students have certain particularities, the psychological condition of the students must be analyzed and reasonably guided. Therefore, the project designed and developed by the author of this article is specifically aimed at colleges and universities. Student HM information data management system.

On the basis of consulting a large number of references related to "HM" and "information management system", this article combines the problems of HM management, K-means clustering analysis and system performance requirements, and analyzes the various aspects of the HM information data management system. The functional module is designed. The system includes 5 functional modules, which are login module, psychological test module, psychological warning module, psychological file management module, and HM forum module. Finally, the performance of the system is tested, and the test results Shows that the performance of the system basically meets the needs of this article.

2 The Design of Mental Health Information Data Management System Under the Background of Informationization

2.1 Problems in Mental Health Management

- (1) The proportion of students who actively seek support is small. According to a survey report from a consulting agency, 23% of students have mental illnesses in need of psychological treatment, 26% of students need psychological counseling, and only 3% of people seek psychological counseling and help. Judging from these statistics, the hidden dangers of students' HM are particularly obvious.
- (2) There are few information systems, and there is often a lack of effective recording and communication channels. At present, most of the psychological counseling institutions set up by colleges and universities do not have effective information methods to provide psychological counseling services for students, and they often maintain face-to-face counseling. Some colleges and universities also use psychological counseling software, which is a tool for recording psychological records,

which is used for preliminary psychological assessment and counseling teachers to record the psychological counseling records of college students. Neither the configuration of software or hardware nor the utilization of college tutoring teachers can be used effectively, which greatly weakens the role of college tutoring agencies.

- (3) Lack of effective counseling tracking for college students receiving psychological counseling. Student psychological counseling is a gradual process, and it is necessary to create an effective file for all students receiving psychological counseling. If there is no information system to store student psychological counseling cases, the results of psychological counseling often have serious deviations.

Information technology can solve these problems. By establishing an information system, you can manage college students' HM consultations online, and at the same time, by using mobile Internet technology, college students' HM consultations can conveniently conduct HM consultations anytime and anywhere through mobile phones, pads and other mobile devices in real time.

2.2 K-means Cluster Analysis

The iterative optimization of the K-means algorithm process is to extract the most valuable part from a large amount of stored psychological data of students. By establishing a psychological data model to provide guidance to managers, the psychological process of students can be managed scientifically, so that it can be improved. Efficiently improve the psychological counseling of students, and the overall effectiveness of the management system can serve as an early warning of risk factors.

In the clustering method, the frequently used quantitative methods are as follows:

$$J_C = \sum_{j=1}^C \sum_{k=1}^n \|x_k^{(j)} - m_j\|^2 \quad (1)$$

$$m_j = 1/n_j \left(\sum_{j=1}^n x_j \right) j = 1, 2, \dots, c \quad (2)$$

In the formula, m_j represents the mean value of samples in the j -th category, and n_j represents the number of samples in the j -th category. The larger the value of J_C , the larger the clustering error. On the contrary, if the value of J_C is smaller, the algorithm has a good clustering effect.

2.3 System Performance Requirements

(1) Reliability

When analyzing the reliability requirements of the system according to the characteristics of the system, this article mainly focuses on the stability of the system in actual operation, such as the ability of the system to withstand failures and the ability to recover. If the

software system is relatively mature, then the system should be stable, able to detect system failures and reduce the possibility of failures, and the system should also have good self-healing capabilities. When analyzing the reliability requirements of the HM data management system in the context of informationization, the main system indicators considered are the system's use time of 5 years, the frequency of failures once a year, the maximum failure of 6 h, the duration and the self-healing time of the system. It can be up to 4 h long.

(2) Security

When analyzing system security requirements, it is necessary to consider the characteristics of the large user group and massive data of the HM data management system under the background of informationization, and analyze from the two aspects of system user security and data. When it comes to system user security, we must first distinguish between user login methods and user password management to better protect system user information. If the user-defined password is relatively simple, the system will automatically request to increase the complexity of the password; from the perspective of system data security consider that you need to define different database permissions for users with different permissions. If the permissions and roles of two users are different, the data tables that can be accessed will also be different. Try to avoid accessing all system data.

(3) Ease of use

In the analysis of system ease of use requirements, the main focus is on interface design, user interaction design, ease of operation, and instructions on how to use the system for HM data management. These ease of use requirements are the basic requirements of the software system. Good usability makes it impossible to attract users and improve market competitiveness. For the HM data management system, when designing and developing the system, it is necessary to ensure that the system interface and audio-visual effects can meet the needs of ordinary users, so that users are visually comfortable during use; at the same time, it is necessary to ensure that the user interacts with the system smoothly and comprehensively. To provide users with a good interactive experience; in addition, the operation of the system should be simple and easy to learn, and a detailed system operation manual should be provided to make the system easier for users to use.

3 Experiment

3.1 Login Module

The first step to use the system is to register. The user enters the user name and password on the system registration page, and then clicks the register button. Then click the register button to start the JavaSwing event listener. When the event is triggered, the system establishes a connection with the back-end database through JDBC connection technology, and compares the entries entered by the user with the entries in the login database one by one. Once from the database If you find the same entry, you will know

whether the user is legally used and brought into the program interface from the system. Once the same entry cannot be found in the database, the user is reminded to enter his account and password again. The design flow chart of the login module is shown in Fig. 1.

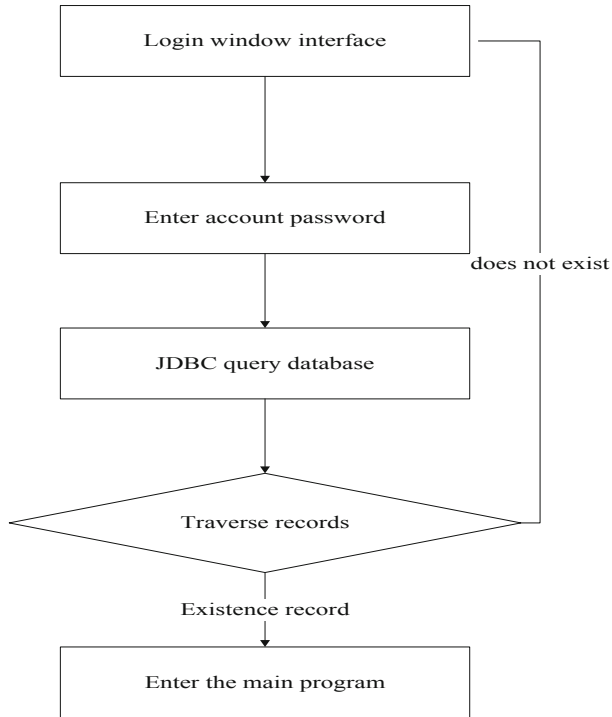


Fig. 1. Flow chart of login module

3.2 Psychological Test Module

The psychological assessment module is mainly based on the psychological test scale for measurement. The user selects a psychological evaluation scale for each HM category, completes the evaluation test questions of each evaluation scale, and submits the test questions. After submitting the test questions, the system will calculate the score based on each standard. According to each score, the corresponding information of the psychological evaluation result will be fed back.

3.3 Psychological Early Warning Module

The psychological early warning module is mainly used by psychological experts to issue psychological alarms based on the user's psychological state, so as to effectively supervise the user's psychological development and early warning of possible risks. After the psychological counselor reviews the results of the psychological test, a psychological warning message is sent to users who are prone to psychological problems. This modular function allows users to more accurately grasp their own HM and find potential problems in time. The psychological counselor can also better grasp the user's HM development and solve it effectively.

3.4 Psychological File Management Module

The function of psychological file is mainly to record the basic situation and key information of students' psychology, so as to better manage the development of students' physical and HM, and guide students' physical and HM. The part of HM file management mainly includes the functions of establishing, testing, editing and deleting HM files. Because only school administrators have the right to establish and delete HM files, HM counseling teachers can grasp the HM of students from the results of school HM tests and psychological counseling research, so as to correctly grasp and establish HM files.

3.5 Mental Health Forum Module

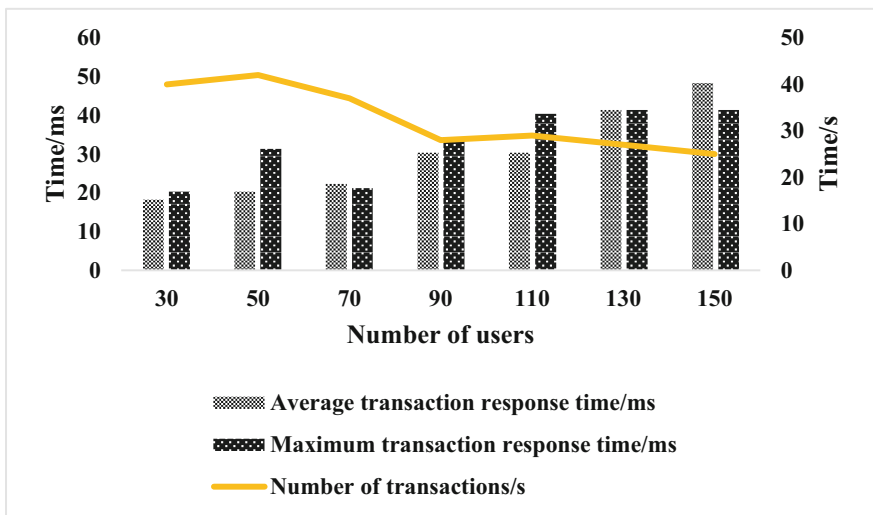
The main page of the system automatically recognizes and fills in the corresponding operation interface according to the visitor ID. Forum users can be divided into administrators, moderators, and ordinary members. Different types of users have different operation permissions. Moderators can keep the relevant content of the entire forum, moderators can only keep the content of the modules with their respective management permissions, members can log in to the forum to view and post, and anonymous users can only view it.

4 Discussion

Performance testing is usually divided into stress testing and load testing. By observing the behavior of the automated testing tool to simulate the program in the real operating environment, it is tested whether the system meets certain development goals. The stress test is used to check the maximum service capacity that the system can provide. Performance control is to ensure that the software can run smoothly and stably under high pressure and heavy load. This test uses Loadrunner software to test the system. The number of concurrent users and transaction execution test cases and test results are as follows. The test example is shown in Table 1.

Table 1. Test cases for the number of concurrent users and transaction execution

Number of concurrent users	Average transaction response time/ms	Maximum transaction response time/ms	Number of transactions/s	Transaction success rate
30	18	20	40	100%
50	20	31	42	100%
70	22	21	37	100%
90	30	34	28	100%
110	30	40	29	97%
130	41	41	27	96%
150	48	41	25	91%

**Fig. 2.** Test cases for the number of concurrent users and transaction execution

It can be seen from Fig. 2 that the basic execution of the transaction will gradually decrease with the increase of the current online user volume. The reason is that the increase in the number of users will occupy more system resources, which will slow down the response time and processing speed of the system. When the number of simulated concurrent users increases to 150, the average transaction response time is 48 ms. At this time, there is no significant delay in actual operation, which is within an acceptable range.

5 Conclusions

Informatization is changing daily life, and school education needs to pay close attention to the development trend of informatization. In the context of information technology, using the most advanced information technology to create a HM information data management system is just in line with the needs of the development of the times. The HM information data management system under the background of information technology transcends the decentralization and concentration of time and place boundaries, makes more reasonable planning and integration of system workflows, and gives full play to the huge advantages of the network in information transmission.

References

1. Gan, X., Yong, L.I.: Application of big data in agricultural informationization in shandong province under the background of "Internet +." *Asian Agric. Res.* **11**(04), 5–7 (2019)
2. Maimunah, M., Haris, H., Priliasari, N.: The design of web-based training management information systems at Pt. Sintech Berkah Abadi. *ADI J. Recent Innov. (AJRI)* **2**(2), 269–274 (2020)
3. Putra, F., Irwanto, T.J., Heryadi, A.Y.: The design and implementation of management information system on student real work (KKN) in Madura University. *Int. J. Civil Eng. Technol.* **10**(2), 159–175 (2019)
4. Choi, J., Taylor, B., Fiszdon, J., et al.: The Impact of mental health services on arrests of offenders with a serious mental illness. *Law Hum. Behav.* **36**(Supplement_2), 170–176 (2019)
5. Burns, T.: International differences in home treatment for mental health problems. Results of a systematic review. *Br. J. Psychiatry* **181**(5), 375–382 (2018)
6. Rutter, M.: How the environment affects mental health. *Br. J. Psychiatry* **186**(186), 4–6 (2018)
7. Thornicroft, G., Semrau, M.: Mental health capacity building in low and middle income countries: the Emerald Programme. *Epidemiol. Psychiatr. Sci.* **27**(1), 1–2 (2018)
8. Davis, K., Coleman, J., Adams, M., et al.: Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *Bjpsych. Open* **4**(03), 83–90 (2018)
9. Dzhambov, A., Hartig, T., Markevych, I., et al.: Urban residential greenspace and mental health in youth: different approaches to testing multiple pathways yield different conclusions. *Environ. Res.* **2018**(Jan), 47–59 (2018)
10. Clark, D.M., Canvin, L., Green, J., et al.: Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *Lancet* **391**(10121), 679–686 (2018)
11. Keetharuth, A.D., Brazier, J., Connell, J., et al.: Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *Br. J. Psychiatry* **212**(01), 42–49 (2018)
12. Dee, T.S., Sievertsen, H.H.: The gift of time school starting age and mental health. *Health Econ.* **27**(5), 781–802 (2018)



Design of Rock Climbing Data Acquisition System Based on LoRa

Dan Zhang¹, Jing Huang²(✉), Luyuan Cai³, and Shuai Cao¹

¹ School of Sports and Leisure, Sichuan Tourism University, Chengdu, Sichuan, China

² School of Physical Education, China University of Geosciences, Wuhan, Hubei, China

³ Mountaineering Management Center, General Administration of Sport of China, Beijing, China

Abstract. As a new sport, rock climbing is deeply loved by people, but it is difficult to intuitively understand the exercise intensity of rock climbers during training. Through the combination of Internet of Things technology and sports, it is possible to monitor rock climbing strength data in real time, and through subsequent analysis of the data, it can help coaches and rock climbers to improve the scientific nature of sports competitive strategies. The main control module of the LoRa-based sports data acquisition system is STM32. The ratio method can be used to calculate the digital signal value of the pressure sensor. The corresponding algorithm is written by establishing the corresponding relationship between the digital signal and the actual strength. Finally, the obtained data is transmitted to the LoRa module store in the computer. Experiments show that the system can provide coaches and rock climbers with accurate training data, provide technical support for coaches to analyze rock climbing technical movements, and improve sports training efficiency.

Keywords: Rock climbing · Internet of Things · STM32 · AD7791 · LoRa

1 Introduction

Rock climbing is an upward climbing sport on a natural or artificial rock wall, and is usually classified as an extreme sport. Rock climbing requires people to continuously complete thrilling actions such as turning, pulling up, moving and even jumping on the rock walls of various heights and different angles. It integrates fitness, entertainment, and competition. It is called “ballet on the cliff”. In 2016, the International Olympic Committee confirmed that rock climbing has become an official event of the 2020 Tokyo Olympics. With the changes in people’s lifestyles, rock climbing is becoming more and more popular in China [1]. However, while people enjoy the fun of rock climbing, their pursuit of technical movements is becoming more perfect and detailed. The combination of Internet of Things technology and rock climbing makes rock climbing becomes intelligent, which can more effectively improve the athlete’s athletic level.

The LoRa-based rock climbing data acquisition system developed this time has certain guiding significance for coaches to detect athletes’ movement data and improve technical movements.

2 System Scheme Design

In order to enhance the maintainability of the system, the system is designed with modular design ideas. Each module is combined with the main control chip to realize the functions of data measurement, storage and real-time transmission [2].

The content to be studied in this article is the design of a rock climbing data acquisition system based on LoRa. The system is mainly divided into two parts: the upper computer and the lower computer. The lower computer is designed on the Cortex-M3 of ARM architecture, combined with the force sensor, front-end amplifier, analog-to-digital conversion component, and wireless transmission LoRa module to realize the collection and transmission of rock climbing data [3]. The upper computer of the computer stores the movement data transmitted by the lower computer. The force sensor built by the bridge and the ADA4528-1 form a front-end amplifier circuit. The amplified signal is sent to the MCU through the analog-to-digital converter AD7791 for calculation to obtain the actual force value, and finally sent to the computer for storage through the IoT kit Lo Ra. The system software simulation is shown in Fig. 1, and the system hardware block diagram is shown in Fig. 2.

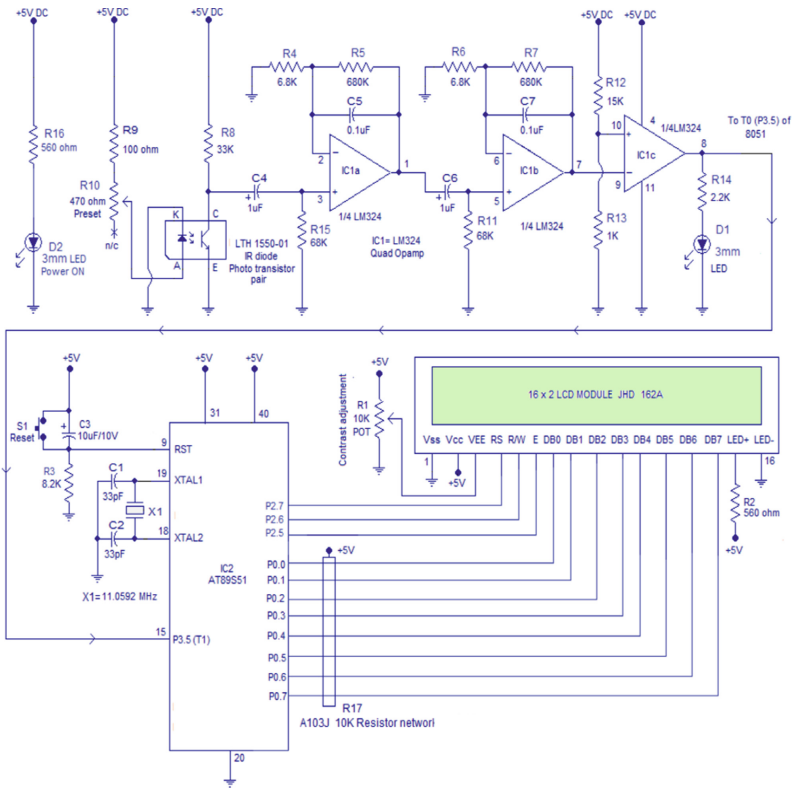


Fig. 1. System software simulation

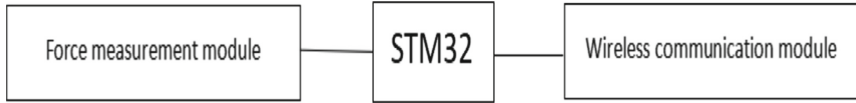


Fig. 2. System hardware block diagram

3 System Hardware Design

The lower computer of the system uses the STM32 chip as the core processor, and the external is composed of a force measurement module and a wireless communication module [4]. The force measurement module mainly includes a force sensor, a differential amplifier circuit, and AD7791 analog-to-digital conversion part.

3.1 Front-End Amplifier Circuit Design

The front-end amplifier circuit is composed of a force sensor composed of a bridge and an external zero-drift amplifier ADA4528-1. ADA4528-1 is a zero-drift operational amplifier especially suitable for amplifying DC and low-frequency low-level signals [5]. It can reduce interference in the signal as much as possible and has rail-to-rail input and output swings. Two pieces of ADA4528-1 and corresponding filters are used to form a differential amplifier circuit, and the reference signal is connected when inputting the signal to avoid the influence of common mode noise. After repeated filtering by multiple filters, the performance index can be well achieved [6]. The capacitors C1, C2, R5, and R6 form a low-pass filter, capacitors C5, R7, and R8 form a differential filter, and C3, C4, R7, and R8 form a common-mode filter. Three types of filters can effectively filter and suppress the amplified signal, and the analog signal after amplification and filtering can be transmitted to the AD7791. The gain calculation formula of ADA4528-1 is:

$$\text{Gain} = 1 + 2 \times R5/R9 \quad (1)$$

This time, the value of front-end R5 is set to 11.3 k Ω , and the value of R9 is set to 60.4 Ω . The magnification factor of ADA4528-1 is 375 calculated by formula (1).

3.2 Design of AD7791 Analog-To-Digital Conversion Module

AD7791 is a 24-bit high-precision analog-to-digital converter using Σ - Δ conversion technology. It selects SPI interface and MCU for data transmission. AD7791 has the characteristics of high accuracy and low power consumption [7]. AD7791 is friendly to low frequency signals and has higher accuracy. AD7791 uses sum-to-difference conversion technology when performing analog-to-digital conversion inside AD7791, which can minimize the impact of noise. Figure 3 shows the A/D conversion timing diagram. The AD7791 measurement formula is:

$$\frac{D}{2^{24}} = \frac{V_1}{Q_i} \quad (2)$$

D is the data after A/D conversion; V1 is the analog signal voltage value obtained by the actual force sensor; S is the AD7791 reference voltage. The ratio method can measure small changes in the signal very well.

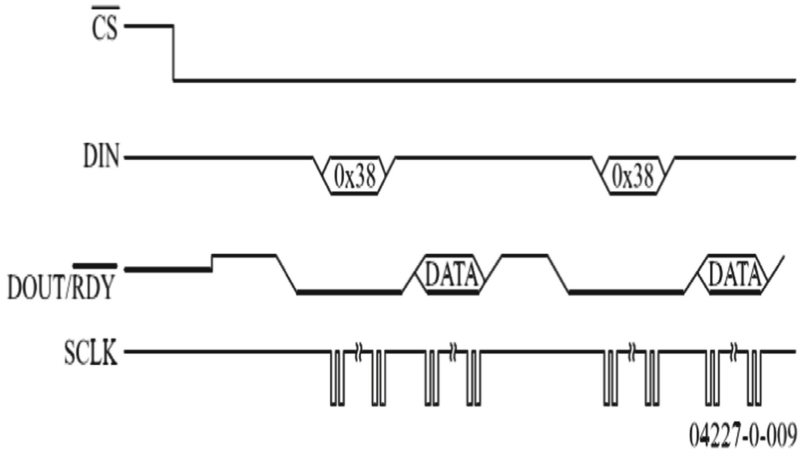


Fig. 3. A/D conversion timing diagram

3.3 Design of Wireless Communication LoRa Module SX1278

LoRa has the characteristics of long communication distance, low power consumption, convenient networking, and good real-time performance [8–10]. The SX1278 chip was selected for this topic. The SX1278 chip can support many modulation modes, including binary frequency shift keying, Gaussian frequency shift keying, binary amplitude keying, and Gaussian filter minimum frequency shift keying. SX1278 has 6 kinds of spreading factors, the frequency range is 7.8 ~ 500 kHz. Figure 4 shows the LoRa block diagram.

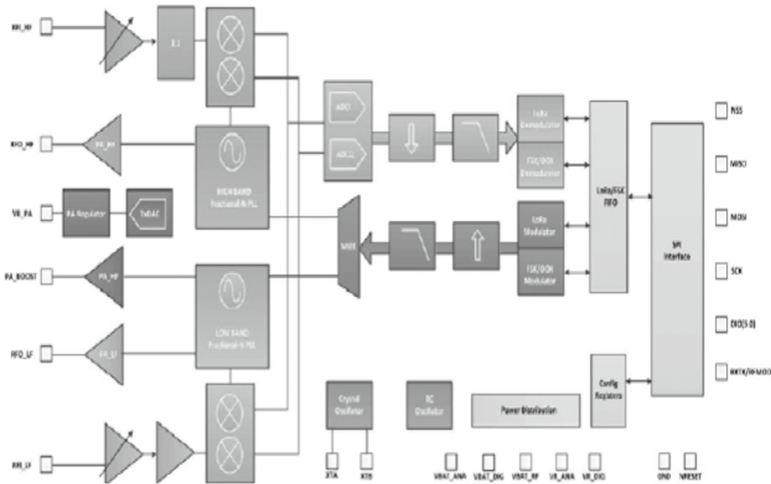


Fig. 4. LoRa block diagram

4 System Software Design

The lower computer software of this system includes GPIO initialization, USART initialization, SPI interface initialization, collecting data, and sending data. Figure 5 shows the software flow of the lower computer.

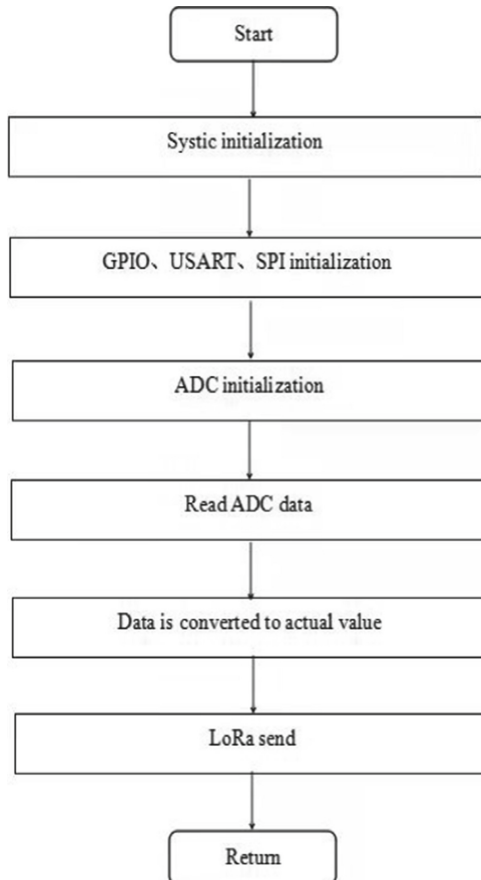


Fig. 5. Lower computer software flow

5 Testing and Analysis

Figure 6 shows the data sampling diagram. The solid line represents the value of the digital signal output by the AD7791 collected within 30 s. It can be seen from the figure that the data fluctuates around 14 500 000, and the dashed line is the data value after the error is eliminated. The value fluctuates around 203 200 000. According to the magnification and analog-to-digital conversion formula, the bridge voltage corresponding to 14 500 000 is 11.523 mV.

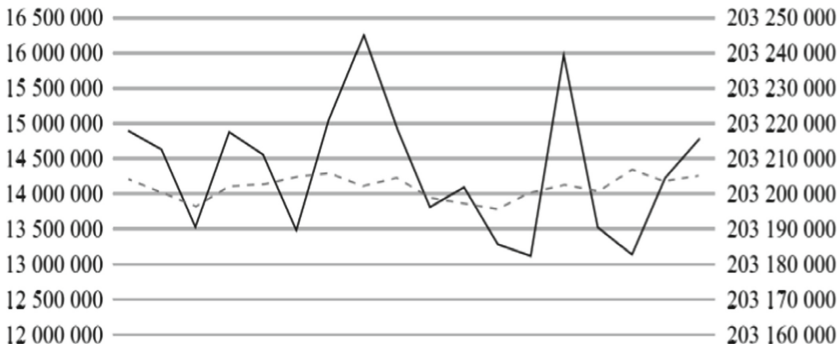


Fig. 6. Data sampling graph

The resistors used in this system are $1\,500\,000\ \Omega$ and $1\,501\,950\ \Omega$. The theoretical value should be $3.242\ \text{mV}$. However, due to the error of the resistor itself, the error shown in the figure above may result in the data. Therefore, when the resistance error reaches $3\ \text{k}\Omega$, The bridge voltage difference has reached $11.562\ \text{mV}$. This study is within the margin of error.

6 Conclusion

Rock climbing has attracted countless enthusiasts with its unique charm. Rock climbers need to have good flexibility, rhythm and climbing skills in order to flash and move easily on the rock wall. Therefore, they have high requirements for action skills. The LoRa-based rock climbing data acquisition system designed in this research has been tested and found that the system runs stably and the error is within the allowable range, so it has better applicability.

References

1. Joseph, S.R., Hlomani, H., Letsholo, K.: Data mining algorithms: an overview. *Neuroscience* **12**(3), 719–43 (2016)
2. Graziano, A.D., Marchetta, V., Cafiso, S.: Structural health monitoring of asphalt pavements using smart sensor networks: a comprehensive review. *J. Traffic Transp. Eng. (English Edition)* **7**(5), 639–651 (2020)
3. Zhang, D.: Research on college student sports training decision support system based on association rule algorithm. *Revista de la Facultad de Ingenieria* **32**(15), 723–726 (2017)
4. Katoh, M., Aihara, T.: Effect of mass of drop-hammer in impact punching with PVA gel as working medium. *J. Jpn. Soc. Technol. Plast.* **60**(697), 27–32 (2019)
5. Xu, L., Jiang, C., Wang, J., et al.: Information security in big data: privacy and data mining. *IEEE Access* **2**(2), 1149–1176 (2017)
6. Chinchuluun, A., Xanthopoulos, P., Tomaino, V., et al.: Data mining techniques in agricultural and environmental sciences. *Int. J. Agric. Environ. Inf. Syst.* **1**(1), 8–12 (2017)
7. Kaspercuk, A., Dardzińska, A.: Comparative evaluation of the different data mining techniques used for the medical database. *Acta Mechanica et Automatica* **10**(3), 233–238 (2016)

8. Guo, Z., Zhang, Z.M., Xing, E.P., et al.: Multimodal data mining in a multimedia database based on structured max margin learning. *ACM Trans. Knowl. Discov. Data* **10**(3), 1–30 (2016)
9. Shoji, H., et al.: Data-mining analysis of the global distribution of soil carbon in observational databases and earth system models. *Geoscientific Model Dev.* **10**(3), 1321–1337 (2017)
10. Perkins, P., Hahn, A., Spratford, W., et al.: Pneumatic boxing glove reduces upward drift in peak force and loading rate over a long series of impacts. *World J. Eng. Technol.* **7**(1), 18–53 (2019)



Research on Realization of Yi Speech Synthesis in Robot

Yihong Jiejue¹(✉) and Saad Metawa²

¹ Liangshan Yi Autonomous Prefecture, Yi Culture Research Center of Xichang University, Xichang 615000, Sichuan, China

² Mansoura University, Mansoura, Egypt

Abstract. As an ancient civilization with a history of more than 5,000 years, our country has formed a multi-ethnic and multi-lingual country through continuous research and progress. It uses multiple languages and scripts in its territory. Yi language belongs to the Han-Tibetan-Burman Yi ethnic group. It is a minority language with a complex and diverse dialect and mother tongue system. According to previous research results, modern Yi language can be divided into six major dialects: north, east, south, west, central and southeast. It includes 25 native languages, showing the diversity and complexity of modern Yi language. The southeast dialect of Yi language can be divided into three kinds of native languages: Sani, Asi and Azhe, This article focuses on the realization of Yi north dialect speech synthesis in robots, and understands speech synthesis technology and Yi speech on the basis of literature data, and elaborates the realization of Yi speech synthesis in robots by designing a Yi speech synthesis system, and test the designed system.

Keywords: Yi Language Speech · Speech synthesis · Synthesis algorithm · System design

1 Introduction

Due to the rapid development of modern advanced science and technology, computer research has promoted the development of artificial intelligence, and human-computer interaction has become very extensive, which has also made computer learning a current research boom [1, 2]. Speech synthesis is to make computers speak, and human-computer interaction systems that use speech synthesis technology are increasingly being studied by many scholars, which also makes many consumer electronics products enter the era of artificial intelligence [3, 4]. This kind of research has great application prospects, whether it is to let the computer make a sound, or let the computer hear the human language. In recent years, language synthesis technology has developed rapidly, especially Chinese has been widely used in various electronic devices. Our country is a large multi-ethnic country, and it is necessary for research to use speech synthesis technology in various languages on computers [5, 6]. The technology of language synthesis is developing day by day, which not only allows computers to produce sounds, but also allows computers

to produce clearer, smoother and more perceptual language effects, which also makes human-computer interaction easier and simpler [7, 8].

Regarding the study of language synthesis, some researchers pointed out that there are very few studies on grouped language. This is because the practice phase of the synthesis model requires more voice samples of a specific speaker to simulate training. At the same time, many natural waveforms are also saved during the synthesis process, and the storage capacity is also quite large. The language prototype of the disguised speech synthesized in this way is derived from the subject's voice, which is more natural than the parameterized synthetic language and has less phase information loss [9]. Some researchers believe that a secret speech recognition algorithm that can use multiple technologies at the same time is more practical than a secret speech recognition algorithm that uses only one specific technology. In practical applications, the voice recognition system will detect secret voices created by various technologies. If the detection algorithm can only detect the speech synthesized by one synthesis technology, the system uses only one built-in detection for each synthesized speech, which not only greatly increases the complexity of the entire system, but also requires more time [10]. Regarding the study of Yi language pronunciation, some scholars proposed that the study of Yi language focused on specific dialects, as well as the distribution and application of Yi language words. These studies are not dedicated to Yi language linguistics, so the materials introduced in some Yi language dialects are not systematic and exhaustive, and even some research methods are not scientific enough, which affects the research on Yi language phonetics, and the one-sidedness in research [11]. In summary, there are more researches on speech synthesis, but relatively few researches on Yi language speech synthesis.

This paper studies the realization of Yi language speech synthesis in robots, analyzes the characteristics of Yi language speech and the basic elements of speech synthesis on the basis of literature data, and then designs a Yi language speech system and carries out the design of the system test, and draw relevant conclusions through the test.

2 Research on Yi Language Phonetics and Speech Synthesis

2.1 The Phonetic Features of Yi Language

- (1) The near sound system of Yi language can include six categories: lips, labial teeth, front tongue, back tongue, uvula, and throat. In terms of vocalization methods, it can be divided into seven-categories: blast, stop, nasal rhyme, side sound, side fricative, fricative, and non-friction [12]. The following explains the Yi language consonant system from these two aspects:

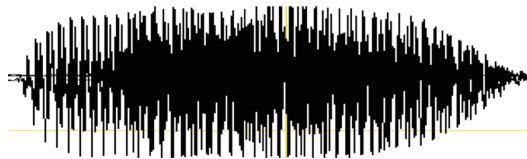


Fig. 1. [t]'s sonogram

1) t

[t] Unaspirated voiceless plosive in front of the tip of the tongue
[t]'s sonogram

2) th

[th] Aspirated voiceless stop before the tip of the tongue
[th]'s Sonogram

- (2) In Yi language, in addition to the consonants directly or by the finals, it may also be composed of the consonants and nasal sounds. “The vowels can be divided into four callings”. The so-called four callings are divided according to the shape of the vowel and the lip shape of the vowel mouth. According to the lip shape, it is divided into four callings: open, Qi, close, and so on.
- (3) Tone refers to the changes in the syllable height of people’s speech, which occupies an important position in social communication and daily activities. Tone is an important feature of Chinese and Tibetan. Some languages have more complex tones, while others have simpler tones. Tone value is a special form of rising and falling changes. The Yi language has four tones: high 55, sub-high 44, middle 33, and low 21.
- (4) Syllable is the basic unit of language structure. The syllable structure is complete and systematic, and it is an important factor for language to form its own language characteristics. Like Chinese, Yi language is a tonal language, and its syllables are usually composed of three parts: consonants, vowels and tones (Figs. 1 and 2).

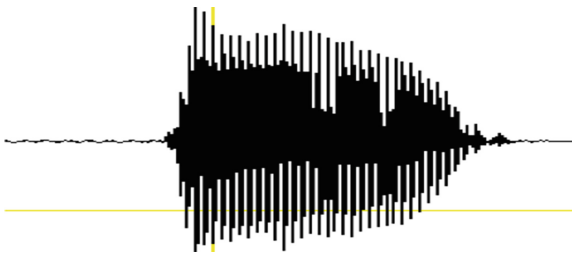


Fig. 2. [th]'s Sonogram

2.2 Basic Elements of Speech Synthesis

(1) Text analysis

To perform speech synthesis, you first need to know what to say, so the first thing in speech synthesis is text analysis. The main function of text analysis is to allow the computer to recognize text from these voices and tell them what to pronounce and how to pronounce. In addition, the computer needs to know what the word is, what the phrase or sentence is, where the pronunciation is paused, and the length in the text.

(2) Prosody generation

Speaking is characterized by rhythm. For example, in Chinese, syllables have different tones, intonations, pauses, and pronunciation lengths. These are all rhythmic functions. Prosody characteristic parameters include acoustic parameters that affect these characteristics, such as fundamental frequency, duration, and volume. The result of the text analysis only shows the pronunciation and the way of pronunciation. This pronunciation is only abstract. For example, if the voice to be expressed in the text is the second or third, and it is emphasized or read a little, where to pause, these are all determined by the rhythm. Similar to the realization of text analysis, prosody can be divided into two methods: rule-based and database-based.

2.3 Speech Synthesis Algorithm

The Gaussian model uses Gaussian probability density function (normal distribution curve) to accurately quantify things, and divides an event into several modules according to the Gaussian probability density function (normal distribution curve). The two main factors that introduce the Gaussian mixture model into speech recognition are: the linear superposition of Gaussian density functions, which can approximate any form of distribution.

In a D-dimensional GMM with M mixed components, it can be represented by the weighted sum of M Gaussian members, namely

$$p(x_t|\lambda) = \sum_{i=1}^M w_i p[x_t|\mu_i, \Sigma_i] \quad (1)$$

Among them, x_t is a D-dimensional observation vector, w_i is the mixed weight, which is equivalent to the probability of each Gaussian member, and $\sum_{i=1}^M w_i = 1$; $p[x_t|\mu_i, \Sigma_i]$ is the D-dimensional Gaussian function, namely

$$p[x_t|\mu_i, \Sigma_i] = \frac{1}{(2\pi)^{D/2} |\Sigma_i|} \exp\left\{-\frac{1}{2}(x_t - \mu_i)^T \Sigma_i^{-1} (x_t - \mu_i)\right\} \quad (2)$$

In the formula, μ_i is the mean vector, and Σ_i is the covariance matrix.

3 Yi Language Speech Synthesis System Design

3.1 System Requirements Analysis

(1) Voice input, processing and analysis of voice database requirements

Before designing the system, you first need to create an audio file database in wave format. Through the above introduction, speech synthesis software can be used to record (analog speech signal and digital speech signal conversion), create speech files, create speech database (only applicable to large-scale speech library creation), if there is a new audio request, the system can add the new audio file to the original database).

(2) Interface requirement analysis

The speech synthesis system provides a graphical user interface, allowing users to work intuitively and conveniently. The main interfaces include:

- 1) Category menu bar (list items and contents according to different categories).
- 2) The content key list (that is, the display of the audio content contained in the selected key, expressed in txt format).
- 3) The function keys include 4 types of controls: input and delete keys, speech synthesis keys, voice listening keys, and file storage keys.

3.2 Text Analysis Module

The text analysis module of the Yi language speech synthesis system processes pronunciation-based text, decomposes any text string into characteristic words, and performs pronunciation marking. At the same time, according to the phonetic and semantic rules, the text analysis part also determines the sentence structure, sound level, sound and various pauses of each word and syllable. The edited text string is converted into a parameter code string that characterizes the pronunciation function. This will be provided for editing later.

- (1) Standardization of input text. Look for spelling errors and exclude irregular or unexpressive characters in the text.
- (2) Word segmentation and morphological analysis. It provides a complete analysis of the vocabulary, grammar and semantics of the text, the division of words and phrases from sentences, and the grammatical relationship between the features of each word and sentence. This is a semantic analysis process. Therefore, the system needs a semantic “dictionary”, and each word in the “dictionary” needs a part-of-speech tag.
- (3) Phonetic transcription. According to the pronunciation, it displays the basic pronunciation unit of each syllable of the sentence, the tone of the sentence, and also analyzes special words (surnames, numbers, proper nouns, special chars, etc.) and polyphony.
- (4) Recognize pitch, pause and accent. Text parsing determines the tone and pronunciation of sentences based on the structure of the text. In addition, in order to resolve the ambiguity in some of the detected accents, it is also necessary to edit the punctuation marks and other characters in the text to determine the pauses between the texts.

3.3 Prosody Generation Module

The prosody part of the text after text analysis, the result of text analysis only tells the system what to pronounce and how to pronounce. This is just an abstract way of pronunciation. The rhythm part determines the tone of the pronunciation, whether it is accented or paused. Similar to the text analysis method, the prosody creation method is divided into two methods: rule-based and data-based.

3.4 Speech Synthesis

First of all, by recording sounds cover all possible accents in the process of human utterance, and then obtain the acoustic parameters of this sound, so as to establish a more complete language library. In the pronunciation process, first select the appropriate acoustic parameters in the audio library for the voice to be expressed, and then generate new voices according to the prosody parameters obtained in the prosody model using synthetic calculations.

4 System Test

4.1 Experimental Design

The experiment in this paper consists of four training sets of four Yi local accents, 35 sentences, 150 sentences, 350 sentences and 550 sentences. In the experiment, this article trained two models, an HD model and a MAT model. The HD model is a related model for Yi dialect speakers, and the MAT model is an adaptive model for Yi speakers.

4.2 Evaluation of Sound Quality

During the evaluation process, 10 Yi local accents reviewers randomly played the speech composition of the HD model and the speech of the MAT model, and evaluated various training sets and two training models. A total of 320 Yi language voice test files (40 Yi language voice test proposals \times 4 language dialect training sets \times 2 models) will be evaluated according to the evaluation criteria. The scoring results are shown in Table 1.

Table 1. Sound quality evaluation results

	HD	MAT
35	1.24	1.76
150	2.34	3.14
350	3.61	3.45
550	4.09	3.89

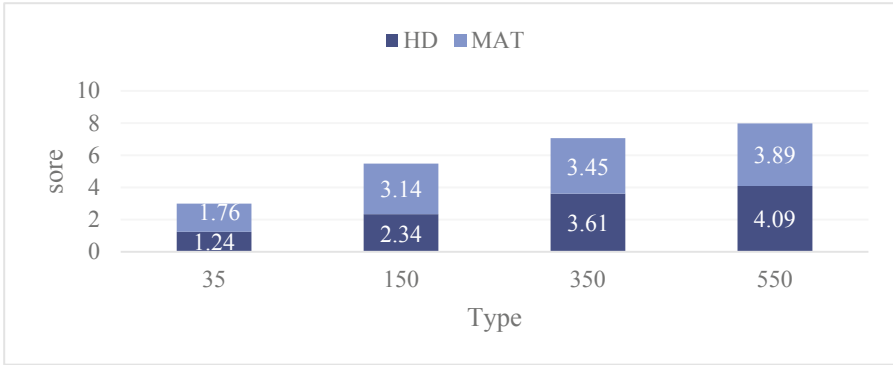


Fig. 3. Sound quality evaluation results

As can be seen from Fig. 3, as the training sentences increase, the scores of the HD and MAT models have improved significantly. Among the 35 proposals, the HD model score is 1.24 and the MAT model score is 1.76. Among 150 audios, the scores of the HD model and the MAT model are 2.34 points and 3.14 points, respectively. In 350 audio, the pronunciation scores of HD model and MAT model are 3.45 and 3.61, respectively. When the proposal reaches 550 proposals, the HD model sounds slightly higher than the MAT model (3.89 and 4.09 respectively). Experiments can prove that the introduction of adaptive dialect training significantly improves the effectiveness of speech synthesis.

4.3 Voice Similarity Evaluation

This article evaluates the similarity between the original Yi language voice and the Yi nationality’s synthesized voice. A total of 320 Yi language audio files (40 Yi language dialect proposals x 4 Yi language dialect training sets x 2 models), play the original voice first, then play the synthesized voice, compare the similarity of the two voices, the evaluation is based on the DMOS score reference score. The results are shown in Table 2.

Table 2. Voice similarity evaluation result

	HD	MAT
35	1.34	2.34
150	2.34	3.23
350	3.45	3.76
550	4.12	4.04

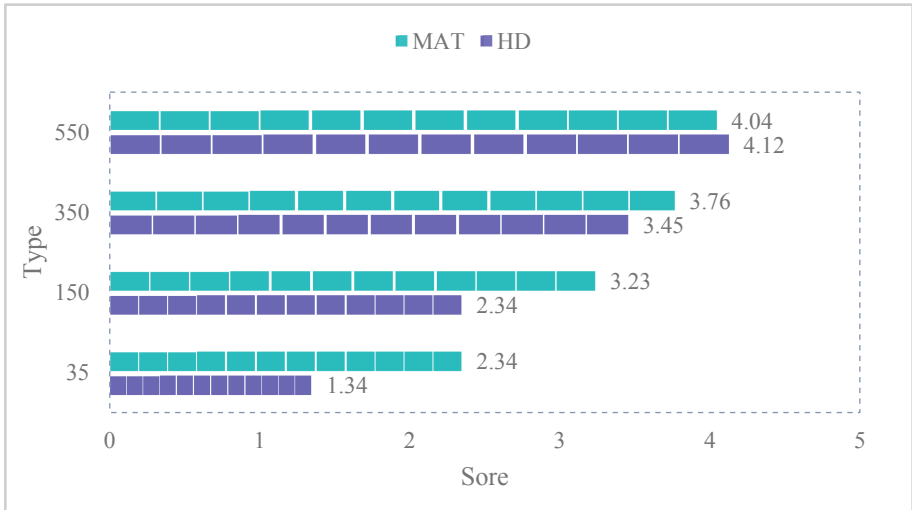


Fig. 4. Voice similarity evaluation result

From Fig. 4 we can see that the HD model scores 2.34 out of 150 sentences and 1.34 out of 35 sentences. Through the comparison, we can see that the similarity has been significantly improved. Similarly, as the number of sentences increases, the speech similarity between the two models will also increase. For the 35-sentence training sentences, the speech composition scores of the HD and MAT models are 1.34 points and 2.34 points, respectively. For the same sentence, the similarity of the MAT model is better than that of the synthetic HD model. In a compound sentence of 550 sentences, the evaluations of the speech composed of the HD model and the speech composed of the MAT model were 4.12 and 4.04 respectively. The audio similarity of HD model is slightly higher than that of MAT. Generally speaking, the speech synthesized by this system has a high degree of similarity. It can be concluded that the more sentences there are, the more similar the synthesized speech is to the original speech. The audio composition similarity of the custom MAT model is better than that of the HD model.

5 Conclusions

This article focuses on the realization of Yi language speech synthesis in robots. After understanding the relevant theories, a Yi language speech synthesis system is designed to illustrate the realization of Yi speech synthesis in robots, and then the designed system according to the experimental test, the experimental results show that the similarity of the speech synthesis of Yi language of the system in this paper is generally high.

References

1. Takamichi, S., Sonobe, R., Mitsui, K., et al.: JSUT and JVS: free Japanese voice corpora for accelerating speech synthesis research. *Acoust. Sci. Technol.* **41**(5), 761–768 (2020)

2. Li, D.: Design of English text-to-speech conversion algorithm based on machine learning. *J. Intell. Fuzzy Syst.* **40**(2), 2433–2444 (2021)
3. Fong, A., Usman, M.: Machine learning for end consumers. *IEEE Consum. Electron. Mag.* **9**(5), 77–78 (2020)
4. Song, E., Soong, F.K., Kang, H.G.: Effective spectral and excitation modeling techniques for LSTM-RNN-based speech synthesis systems. *IEEE/ACM Trans. Audio Speech Lang. Process.* **25**(11), 2152–2161 (2017)
5. Filintisis, P.P., Katsamanis, A., Tsiakoulis, P., et al.: Video-realistic expressive audio-visual speech synthesis for the Greek language. *Speech Commun.* **95**, 137–152 (2017)
6. Rubeena, A.: Concatenative speech synthesis: a review. *Int. J. Comput. Appl.* **136**(3), 1–6 (2016)
7. Narendra, N.P., Rao, K.S.: Generation of creaky voice for improving the quality of HMM-based speech synthesis. *Comput. Speech Lang.* **42**(Mar), 38–58 (2017)
8. Khodabakhsh, A., Mohammadi, A., Demiroglu, C.: Spoofing voice verification systems with statistical speech synthesis using limited adaptation data. *Comput. Speech Lang.* **42**(Mar), 20–37 (2017)
9. Nose, T.: Efficient implementation of global variance compensation for parametric speech synthesis. *IEEE/ACM Trans. Audio Speech Lang. Process.* **24**(10), 1694–1704 (2017)
10. Wang, X., Takaki, S., Yamagishi, J.: Investigation of using continuous representation of various linguistic units in neural network based text-to-speech synthesis. *IEICE Trans. Inform. Syst.* **E99.D**(10), 2471–2480 (2016)
11. Tsiaras, V., Maia, R., Diakouloukas, V., et al.: Global variance in speech synthesis with linear dynamical models. *IEEE Signal Process. Lett.* **23**(8), 1057–1061 (2016)
12. James, J., Balamurali, B.T., Watson, C.I., et al.: Empathetic speech synthesis and testing for healthcare robots. *Int. J. Soc. Robot.* 1–19 (2020)



Application of Deep Learning Algorithm in National Fitness Monitoring

Xiaoliang Li¹ and Liping Zhang²(✉)

¹ Department of Health and HealthCare, Zhangzhou Health Vocational College, Zhangzhou 363000, China

² Department of Public Foundation, Zhangzhou City College, Zhangzhou 363000, China

Abstract. With the continuous improvement of people's living standards, the improvement of material life and the level of science and technology, the people's values and lifestyles have also undergone tremendous changes. This article aims to study the application of deep learning algorithms in national fitness monitoring. Based on the analysis of the current status, problems and causes of the national fitness monitoring work, the questionnaire survey method is used to investigate the cognition, feedback and other issues of the citizens of City S on the fitness monitoring. Propose feasible strategies based on its problems. The survey results show that the citizens themselves have problems with their attitudes, awareness, and cognition when dealing with the work of physical fitness monitoring. However, citizens are still satisfied with the work of physical fitness monitoring and are willing to accept it. Physical fitness monitoring is still popular among citizens.

Keywords: Deep learning algorithm · National fitness monitoring · Monitoring methods · Development strategies

1 Introduction

The physique of the citizens plays a vital role in the development of a country. The development of economy, politics, technology, etc. cannot be separated from the human foundation [1, 2]. With the proposal of “Healthy China” in 2015, the building of a healthy China has been raised to a strategic height [3, 4]. With the development of the National Fitness Program, people have begun to pay more and more attention to the development of physical fitness. National physique monitoring has also become more and more important [5, 6].

My country's national fitness monitoring is relatively late, and in this respect it is still in a relatively immature stage compared with Western countries [7, 8]. Therefore, research on citizen physical fitness monitoring should not only be valued by experts and scholars, but should also receive the attention of the whole society. The quality of the national physique will directly affect the economic development and social stability of a country [9, 10]. Therefore, it is of great significance to understand the status of national fitness monitoring.

Based on the analysis of the status quo, problems and the causes of the problems in the national fitness monitoring work, this paper uses a questionnaire survey method to investigate the cognition and feedback of the citizens of City S on the fitness monitoring, and proposes feasible strategies based on the problems. The survey results show that the citizens themselves have problems with their attitudes, awareness, and cognition when dealing with the work of physical fitness monitoring.

2 Research on National Physique Monitoring Methods Under the Background of Big Data

2.1 Analysis of the Status Quo of Physical Fitness Monitoring

(1) Monitoring network links

The various indexes of national fitness not only reflect the physical condition of the tested object, but also reflect the adaptability of people to the geographical environment to a certain extent. In the future development trend, it is necessary to introduce GIS technology, that is, geographic information system technology, to establish and improve the national fitness assessment model [11, 12].

(2) Links of monitoring tasks

The national fitness monitoring work involves many links, which means that relevant monitoring tasks run through the macro structure and microstructure of the national fitness monitoring network. From the perspective of the operation process of the monitoring task, the big data technology platform can run through the entire process of the national fitness monitoring task, which requires many government departments such as sports, education, health, labor unions, women's federations, science and technology to achieve cross-border integration and digital resource sharing.

2.2 Problems in Physique Monitoring

(1) Monitoring network links

The first flaw is the failure to reflect the urban-rural differences through big data. Air quality, living environment and many other factors make it difficult for the same physical fitness monitoring standard to reflect the true physical status of different monitored people, and the setting of the monitoring network loses objective evaluation standards. The second flaw is that there is no technical support of big data to pay attention to the national physique monitoring of the floating population.

(2) Links of monitoring tasks

In the micro-level operating system, vertical resource sharing and intercommunication in the same field are relatively good, but there are still many barriers to horizontal resource sharing and intercommunication in different fields. Taking the physique monitoring task of the elderly as an example, the coordination and cooperation between the neighborhood office, the elderly service agencies, and the property management of residential communities in the physique monitoring task of the elderly is still at a low level. The national physique monitoring task link there is still a great expansion in the high-level collaboration and cooperation of sharing big data resources.

(3) Sampling link of monitoring objects

In the process of national physique monitoring in City S, the sample sampling did not fully reflect the technical advantages of big data. More importantly, our country's increasingly accelerating social mobility and rapid urbanization process have gradually eliminated the traditional dualistic social pattern, increasing the complexity of monitoring objects to a certain extent, which also requires great support from an objective point of view.

2.3 Analysis of the Cause of the Problem

(1) Analysis of the causes of problems in hardware construction

Big data, as a product of information technology innovation, not only puts forward requirements for related hardware construction, but also requires relevant professional staff to operate. The complexity of national physique monitoring determines that relevant staff must be temporarily recruited, which means that it is difficult to ensure the consistency and standardization of relevant staff in terms of information technology capabilities.

(2) Analysis of the causes of problems in software construction

Relying on big data technology to improve the scientific level of national fitness monitoring is a specific application of information technology to improve people's livelihood. However, the poor health awareness of the broad masses of the people at the grass-roots level in our country or their resistance to government-led cooperation or even confrontational psychology, as well as the lack of awareness of the advantages of the emerging technology of big data by the implementers of the monitoring work, have determined the major disadvantages. The complexity and difficulty of national physique monitoring in the context of data.

(3) Analysis of the causes of problems in the work concept

The emergence of big data is an important change in the changes of human society, which requires timely follow-up of relevant theoretical knowledge. However, the limitation of the theoretical level of our country in the process of improving the national physique determines that it is difficult to form a corresponding theoretical knowledge system in a short time. This objectively means that the national physique monitoring work in the context of big data has weaknesses and deficiencies in the theoretical knowledge system.

2.4 Advantages of Deep Learning Algorithms

(1) Ability to abstract complex features. Deep learning has a deep network structure. It can not only reduce the dimensionality of complex feature data to a certain extent, but also extract deep features, which can effectively use deep features to characterize data.

(2) Feature extraction based on unsupervised pre-training. Traditional machine learning is based on manually extracting features, which not only relies on human experience, but once the feature selection is inappropriate, it will directly lead to a decrease in the accuracy of the supervised learning results.

- (3) Practice shows that deep learning has made significant achievements in the field of image processing and pattern recognition, and its accuracy is better than other algorithms.

3 Experiment

3.1 Questionnaire Survey Method

3.1.1 Formulation of the Questionnaire

After reading relevant materials, books, topics, etc., this article focuses on the national physical fitness monitoring methods, and understands the current status of the physical fitness monitoring work in City S. The questionnaire is formulated and revised on this basis. After the questionnaire design is completed, experts are searched for review and guidance. After reviewing the questionnaire, the experts give many pertinent suggestions and correct the existing problems. Many revisions were made to the questions and suggestions raised by experts to ensure the validity of the questionnaire. The questionnaire design is aimed at some 600 adult citizens in the urban area of S city, selected citizens according to the principle of random stratified sampling.

3.1.2 Reliability Test of the Questionnaire

Using formula (1) to calculate the reliability coefficient, the correlation coefficient of the questionnaire is $r = 0.883$. According to the theories and methods of modern scientific research, when the reliability of a test reaches 0.80 or more, it can be regarded as a test with higher reliability. The test results confirm that the questionnaire is reliable.

$$s^2 = \frac{(M - x_1)^2 + (M - x_2)^2 + (M - x_3)^2 + \dots + (M - x_n)^2}{n} \quad (1)$$

$$r = 1 - S^2(1 - r_1) / S_n^2 \quad (2)$$

$$r = \frac{2r_{ban}}{1 + r_{ban}} \quad (3)$$

3.1.3 Issuance and Collection of Questionnaires

The citizen questionnaire in this article is distributed to 100 citizens randomly selected from each district in six districts and counties of S city. 600 questionnaires were distributed, 584 were returned, and 574 were valid questionnaires. The method adopted is to personally distribute and recycle on-site and others to distribute and recycle on-site. When others substitute for distribution, they will be notified of the precautions to be filled in, and then the agent will convey it to the person who fills in.

Table 1. Understand the degree of fitness testing methods

Degree	Frequency	Proportion
Know it well	62	10.8%
Understanding	164	28.57%
General understanding	246	42.86%
Don't understand	74	12.89%
Don't understand at all	28	4.88%

4 Discussion

4.1 Citizens' Awareness of Physical Fitness Monitoring in S City

Table 1 shows the level of understanding of the 574 citizens surveyed on physical fitness testing methods. There are 62 people who know very well, accounting for 10.8%, 164 people who know, accounting for 28.57% of the total, there are generally 246 people, accounting for 42.86% of the total, 74 people who don't know, accounting for 12.89%, very there are 28 people who don't understand, accounting for 4.88%. From the table, we can clearly see that more than two-thirds of the citizens have some understanding of the physical fitness test methods, and only a small number of citizens do not. It shows that most citizens have participated in the physical fitness test, which shows that City S has put this work into practice and considers the citizens personally.

4.2 Citizens' Feedback on Physical Fitness Monitoring

Table 2. Physical fitness test projects participated by citizens

Monitoring content	Frequency	Proportion
Height, weight, vital capacity, step test, sitting forward bending	498	86.45%
Vertical jump, stand on one foot with closed eyes, select reaction time, push-ups, sit-ups	526	91.31%
Body composition analysis, bone density	352	61.11%
ECG, blood sugar, blood lipids, blood routine, spine measurement	69	11.97%
Other projects	236	3.64%

Table 2 shows the test items that Beijing urban residents came into contact with during the physical fitness test. Most of the citizens have participated in the 11 national test indicators of height, weight, vital capacity, step test, sitting forward bending, grip strength, vertical jump, standing on one foot with closed eyes, selective reaction time, push-ups, and sit-ups. Participated in body composition analysis, and bone density 352 people, accounting for 61.11%. There are 69 people who have participated in electrocardiogram, blood sugar, blood lipid, blood routine, and spine measurement, accounting for 11.97% of the total. Others accounted for 3.64%.

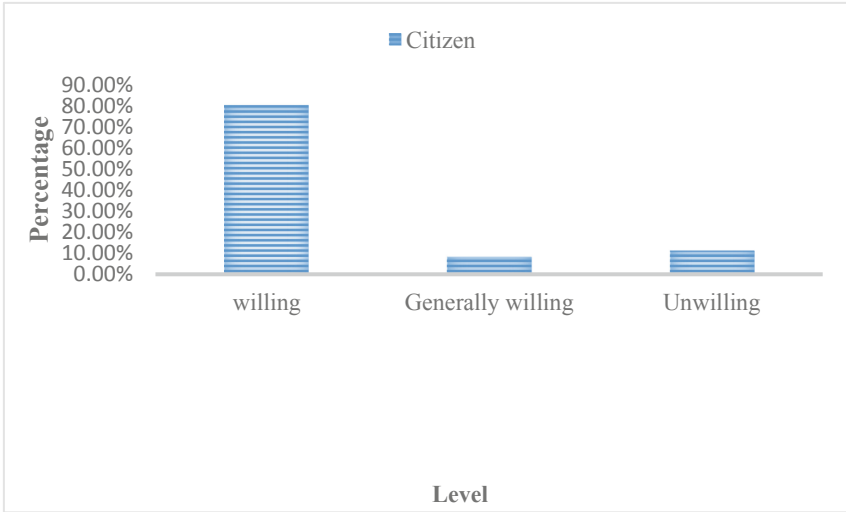


Fig. 1. Are you willing to save the test data

Figure 1 shows whether citizens are willing to save the monitoring data. Among them, 463 people, 80.38% of the total, expressed their willingness to establish their own data accounts for future comparative analysis and real-time inquiries about their physical condition. In general, 48 people, accounting for 8.33%, expressed that they were unwilling to save the monitoring data, and 65 people thought it would involve privacy and felt that it was of little use. On the whole, citizens are still willing to build their own databases.

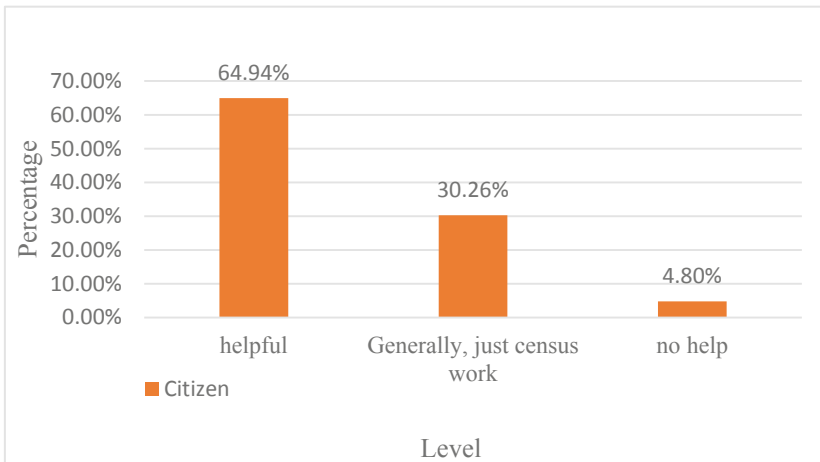


Fig. 2. Does the physical fitness monitoring work help citizens' health?

Figure 2 shows whether citizens think the physical fitness monitoring will help citizens' health. There are 374 citizens who believe that carrying out physical fitness monitoring is very helpful to their own health, can gain an understanding of their physical fitness, and enhance their awareness of physical exercise and the degree of concern for sub-health. Approximately 64.94% of the total percentage. 30.26% of people said they were fair and thought it was just a census. 4.8% said it did not help. Overall, it shows that citizens are still satisfied with the work of physical fitness monitoring and are willing to accept it. Physical fitness monitoring is still popular among citizens.

5 Conclusions

The national physical health condition affects the stable development of a country and a region. Citizens are the mainstay of creating social wealth. Facing the current rapid development of society, the accelerating pace of life, and increasing pressure from all aspects of society, citizens must have a good physical fitness level. Therefore, based on the analysis of the status quo, problems and causes of the problems in the national physical fitness monitoring work, this article uses a questionnaire survey method to investigate the cognition and feedback of the citizens of City S on physical fitness monitoring and other issues, and propose feasible strategies based on the problems.

Acknowledgments. Project information: Economic Research Center of regional reform and Development Institute of Chinese Academy of Management Sciences project "implementation and discussion of national physique monitoring in the era of big data" No.: jjyj6008.

References

1. Beh, P.K., Ganesan, Y., Iranmanesh, M., Foroughi, B.: Using smartwatches for fitness and health monitoring: the UTAUT2 combined with threat appraisal as moderators. *Behav. Inf. Technol.* **40**(3), 282–299 (2021)
2. Wla, D., Ah, B., Ping, W., et al.: PbFG: physique-based fuzzy granular modeling for non-invasive blood glucose monitoring. *Inf. Sci.* **497**, 56–76 (2019)
3. Krimmer, J., Angellier, G., Balleyguier, L., et al.: A cost-effective monitoring technique in particle therapy via uncollimated prompt gamma peak integration. *Appl. Phys. Lett.* **110**(15), 1281 (2017)
4. Gonzálezruiz, G.E., Borréortiz, Y.M., Oduber, E.R., et al.: Prevalencia y factores asociados a osteoporosis en pacientes de Santa Marta (Colombia). *Journal De Physique* **21**(6–7), 561–571 (2016)
5. Folkvord, F., van Breugel, A., de Haan, S., de Wolf, M., de Boer, M., Abeele, M.V.: A protocol study to establish psychological outcomes from the use of wearables for health and fitness monitoring. *Front. Digit. Health* **3**, 708159 (2021)
6. Helms, E.R., Prnjak, K., Linardon, J.: Towards a sustainable nutrition paradigm in physique sport: a narrative review. *Sports* **7**(7), 172 (2019)
7. Salmund, J.A., Tadaki, M., Dickson, M.: Can big data tame a "naughty" world? *Can. Geogr.* **61**(1), 52–63 (2016)
8. Panicacci, S., Giuffrida, G., Donati, M., Lubrano, A., Ruiui, A., Fanucci, L.: Empowering home health monitoring of covid-19 patients with smartwatch position and fitness tracking. In: *CBMS 2021*, pp. 348–353 (2021)

9. Ramakrishnan, R.K., Gavas, R.D., Viraraghavan, V.S., Hissaria, L.K., Pal, A., Balamuralidhar, P.: Monitoring the early physiological effects of fitness regimes using PPG signals. In: PerCom Workshops 2020, pp. 1–5 (2020)
10. Sumit, M., Tapas, M., Deen, M.: Wearable sensors for remote health monitoring. *Sensors* **17**(12), 130 (2017)
11. Qing, X.P., et al.: An active diagnostic system for structural health monitoring of rocket engines. *J. Intell. Mater. Syst. Struct.* **17**(7), 619–628 (2016)
12. Wachowicz, A., Małysiak-Mrozek, B., Mrozek, D.: Combining data from fitness trackers with meteorological sensor measurements for enhanced monitoring of sports performance. In: Rodrigues, J.M.F., et al. (eds.) ICCS 2019. LNCS, vol. 11538, pp. 692–705. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22744-9_54



The Optimization Algorithm of Tourism Route Based on Deep Learning Theory

Jia Xue^(✉)

Chengdu Polytechnic, Chengdu 610041, Sichuan, China

Abstract. With the rapid development of the tourism industry, the traditional fixed single tourism route can no longer meet the needs of consumers. The design of a reasonable, individualized and diversified tourism route has become an urgent problem for the tourism industry. The design of tourist routes is mainly divided into tourist node selection and node combination optimization. This paper designs a travel route based on deep learning theory and uses intelligent algorithms to optimize the route. This article first uses the literature review method and the case verification method to explain the concept and characteristics of creative tourism; secondly, it describes and introduces the background of the new era, in which computer technology accounts for a large proportion; finally, it also constructs the integration In order to select more suitable indicators, a measurement indicator system was constructed to test the level of integration of creative industries and tourism industries. This article takes two more famous tourist cities in Jiangxi Province and draws the following conclusions: In 2019, Nanchang achieved an added value of tourism of 149.6 billion yuan, an increase of 5% over the previous year, accounting for 7.5% of the city's GDP of 12.4%. Tourism has gradually become the economic support of Nanchang!

Keywords: Deep learning theory · Travel route · Optimization algorithm · Route design

1 Introduction

In any corner of the world, different industries are actively or passively exploring and developing new concepts and new models, exploring the possibility of interaction and integration between industries. Tourism is a comprehensive industry and an important driving force for economic development. In recent years, with the improvement of comprehensive national strength, the people have no worries about food and clothing. Tourism has developed rapidly in recent years and has become a hot spot for residents' consumption.

Regarding the tourism industry, various experts have also expressed many opinions. Martín believes that cultural and creative tourism is regarded as the two main industry support for tourism and tourism groups, including direct support and services provided by tourism support groups and software industry groups; tourism industry groups include travel agencies, hotels, tourism and entertainment, Tourist market and tourist transportation. It belongs to the tourism industry included in the primary and secondary industries,

as well as other related industries in the tertiary industry besides tourism [1]. Richards not only combines cultural and economic factors, but also connects with the background of the current era. He summarizes the cultural creative industry as “in the era of global consumption, a new economic form that relies on high technology and the Internet to satisfy people’s spiritual and cultural aspects” [2]. Based on the research foundation of American scholars, Lopes also proposed the “5Cs” structure for the development of Hong Kong’s cultural and creative industries. His argument is based on the perspective of creative capital. His basic point is: the creative industry includes five types of capital, namely human capital and creativity [3].

This article focuses on new forms, based on the richness of existing tourism resources, takes cultural creativity as the theme of tourism development under the background of the new era, takes tourism+ as the main line for tourism development, and seeks the theory of the integration of cultural and creative industries and traditional tourism. Tourism industry development machine [4].

2 Method

2.1 The Concept of Deep Learning Theory

By constructing a multi-level intelligent learning model and massive training data, it extracts the essential characteristics of the data and analyzes the potential value of big data to solve complex abstractions. The problem. Deep learning has made a major breakthrough in the research field of traditional neural networks. First, deep learning can achieve hierarchical expression of data and complex function approximation functions by constructing deep nonlinear networks, and characterize the essential attributes of data, solving complex abstractions Data visualization and classification, prediction and other issues; second, the feature learning process of deep learning is an unsupervised nature, through a top-down unsupervised training to learn data features, and then apply top-down Supervised training is used to fine-tune the parameters of a network. The comprehensive training of “layer-by-layer pre-training” and “fine-tuning” solves the problem of neural network parameter adjustment.

2.2 Analysis of “Tourism+”

- (1) In the context of the new era, the development of “tourism+” does not seem to be unexpected. Integrating regional characteristic culture into cultural and creative products and strengthening the connection with modern life will help tourists and the public who have a sense of identity with local culture. In modern tourism, people are no longer confined to the tourism form of walking horses and flowers, but also hope to understand the local material and spiritual culture during the travel process, so that more people can enjoy the baptism of local culture [5].
- (2) In the tourism era, tourism-related industries are quietly merging, such as “tourism + cultural creativity + agriculture”, “tourism”, “cultural creativity” and “network”. As a product of tourism cultural creativity in the tourism industry, the power of integration can connect related industries, lead the development of local industries, and spread local culture to the society, forming a “mutual integration and mutual assistance in tourism, culture, network, economy, and agriculture” [6].

2.3 Features of Creative Tourism

Creative tourism has changed the traditional way of tourism based on projection, seeking the active participation and learning of tourists, and meeting the needs of tourists for individualization and self-improvement. As a product of the integrated development of the creative industry and the tourism industry, creative tourism has its own unique characteristics:

- 1) Culture is the soul of tourism and the foundation for the survival and development of tourism. The creative industry is a product of social, economic and cultural development to a certain extent, and is the product consumption of people's growing knowledge and cultural needs. As a new type of activity combining creative industry and tourism industry, creative tourism has realized the transformation from culture to creativity [7].
- 2) Creativity is an important feature that distinguishes creative tourism from other forms of tourism. The extensive development model of the traditional tourism industry not only requires a large investment of human, material and economic resources, but it also causes great damage and waste to the environment and tourism resources. Creative tourism is no longer a passive sightseeing tour. Tourists must invest in various elements such as wisdom, and participate in the feelings and skills of tourism creation during the travel process.
- 3) Interactive interactive participation is a way to realize creative tourism. From the above concept of creative tourism, creative tourism emphasizes the "participation in activities" and "learning experience" of tourists. The biggest feature different from traditional tourism is that the acquisition of creative experience not only depends on the provision of tourist destinations, but also requires tourists to participate in the creation. Only tourism activities that combine with tourists and destinations are truly creative.
- 4) In general tourism activities, tourism resources (things) dominate the development of tourism activities or projects. Creative tourism is the crystallization of human wisdom. It mainly relies on human subjective resources, such as creativity, wisdom and skills. These intangible spiritual resources are inexhaustible, inexhaustible, and can be recycled forever, so he said that creative tourism has obvious sustainability [8].

2.4 Creative Tourism and Cultural Tourism

The relationship between tourism and culture has always been an important content of tourism research. Creative tourism and cultural tourism are closely related, but there are differences between the two. From the perspective of the definition and basic elements of creative tourism, creative tourism and cultural tourism are essentially the same, that is, the essence of creative tourism and cultural tourism are the same. Culture is considered to be the main content of tourism products, but from the analysis of content and characteristics, cultural tourism pays more attention to the cultural concept of tourism products and pay more attention to culture [9].

2.5 Fusion Model

The degree of interaction and interconnection between systems is called the degree of coupling. In the empirical analysis of this article, the degree of coupling can be used to measure the result and degree of the interaction between the cultural creative industry and the tourism industry. The integration model is widely used in the measurement of industrial integration. This paper chooses more suitable indicators to construct a measurement indicator system to test the level of integration of creative industries and tourism industries [10]. Set variable $\mu_i (i = 1, 2, 3, 4, \dots, n)$ Represents the system. When $n \geq 2$, the coupling degree model of the interaction between n systems is:

$$C_n = n \left\{ \frac{\mu_1 \times \mu_2 \times \dots \times \mu_n}{(\mu_1 + \mu_2 + \dots + \mu_n)^n} \right\}^{1/n} \tag{1}$$

Let $n = 2$, the coupling degree model becomes:

$$C_2 = n \left\{ \frac{\mu_1 \times \mu_2}{(\mu_1 + \mu_2)^2} \right\}^{1/2} \tag{2}$$

This formula can be used to measure the degree of coordination between cultural and creative industries and tourism in the process of industrial integration among them, $\therefore C_2$ larger, the higher the coupling.

3 Experiment

3.1 Literature Research

The classification and further summary of related documents constitute the theoretical basis of this article on the integration and development of creative industries and traditional tourism. At the same time, it summarized recent policy documents and program texts to determine the social and economic background for achieving this goal.

3.2 Logic Deduction

This article first defines the relevant research concepts, and then summarizes the elements of the creative cultural industry and tourism industry system. Secondly, it summarizes the integration characteristics of cultural creative industry and tourism industry, and constructs a measurement model of the integration level of cultural creative industry and tourism industry.

3.3 Combination of Theory and Practice

As a new form of industry, it is impossible to comprehensively analyze the tourism creative industry through a single theoretical case study and analysis. Based on the industrial characteristics of the tourism creative industry, this article combines relevant industry integration theories and classic cases of tourism creative industry development to further analyze and guide the direction and trend of the development of industry integration.

4 Discussion

4.1 Status of Jingdezhen Tourism and Cultural Industry

The tourism industry in Jingdezhen has developed into a pillar industry, with tourism revenue accounting for 35% of the region's GDP in 2019. Tourism is listed as one of the future "3 + 1" characteristic advantageous industries in the region. (Ceramic industry, aviation industry, automobile industry and tourism industry).

From Fig. 1, Jingdezhen's tourism revenue accounted for one-third of the annual GDP, and its cultural and creative industries have a solid foundation. The cultural and creative industry is the foundation and soul of cultural and creative tourism. Cultivate and practice more than 3,000 ceramic cultural and creative industry operators, of which more than 12 have been identified as provincial-level cultural industry demonstration bases, and three have been included in the province's cultural and creative service industry complex.

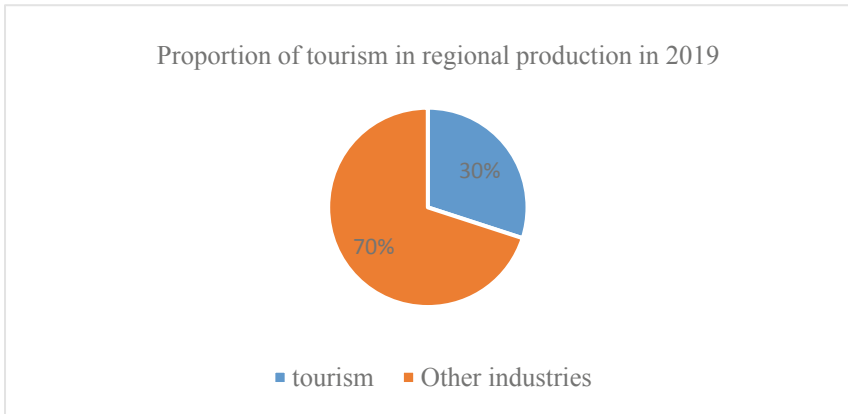


Fig. 1. The proportion of tourism in GDP in 2019

4.2 The Development of Cultural and Creative Tourism-Related Industries is Steady-Taking Nanchang as an Example

The overall development of Nanchang's tourism industry is in a steady state of development. During the "Eleventh Five-Year Plan" period, Nanchang's tourism industry developed vigorously. Through continuous integration of tourism resources and actively building regional tourism brands, the status of Nanchang's tourism industry continued to rise, the scale of the industry continued to expand, and tourism consumption continued to increase (Table 1).

Analyze from the perspective of the growth value of the tourism industry in Nanchang. During the "Eleventh Five-Year Plan" period, the added value of Nanchang's tourism industry grew at an average annual rate of 26.58%, higher than the growth

Table 1. Nanchang tourism growth value table

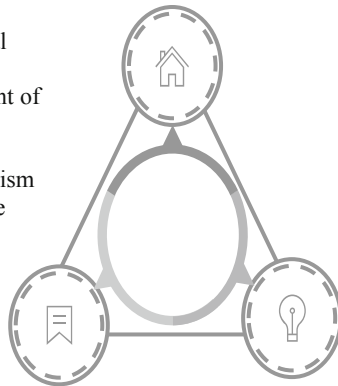
Index	2010	2015	2018	2019
Added value (100 million yuan)	500	1300	1411	1496
Proportion of added value of tourism industry in Nanchang’s GDP (%)	6.2	8	7.4	7.5

rate of the tertiary industry in Nanchang during the same period. In 2015, Nanchang’s tourism industry achieved an added value of 130 billion yuan, an increase of 133% over the end of 2010, accounting for 10% of the added value of Nanchang’s tertiary industry during the same period and 8% of Nanchang’s GDP during the same period. In 2019, Nanchang achieved an added value of 149.6 billion yuan in tourism, an increase of 5% over the previous year, and accounted for 12.4% of the 7.5% of the city’s GDP. Tourism has gradually developed into a pillar industry in Nanchang.

4.3 Fusion Mechanism

Based on the perspective of system theory, the concept of industry is defined as a systematic combination of technology, product, enterprise, market and system. Different industrial factors interact with each other, continuously differentiate and restructure, and promote the development of the industrial system. It can be said that the integration mechanism of industries is the result of the evolution of different divisions of labor between or within industries (Fig. 2).

Culture: the concept of industrial integration
 Technology: technological innovation
 Economy: the development of corporate groups
 Politics: government deregulation, cultural tourism
 Policy support for creative industries



The pursuit of interests by enterprises and the improvement of industrial value chain structure
 Continuous innovation in the view of tourism resources
 Creative talent

Tourism market demand

Fig. 2. The integration mechanism of tourism industry and cultural creative industry

4.4 Enhance Brand Awareness

If a city wants the tourism industry to develop steadily, it must strengthen brand investment, improve the planning, marketing, and promotion capabilities of park

brands, industry brands, and corporate brands, and increase the visibility and reputation of the cultural industry. Implement enterprise growth projects. Speed up the cultivation of a number of cultural brands with good development prospects, excellent business performance, high industry visibility, high market share, and independent intellectual property rights. Enterprises should develop in the direction of specialization, precision, speciality, and new, and use capital as the link to carry out cross-regional, cross-industry, and cross-ownership mergers and reorganizations. Raise funds from the capital market, cultivate a group of strategic investors in the cultural field, and further expand and strengthen.

5 Conclusion

Based on the research of experts and scholars, this article analyzes the cultural and creative tourism industry. It discusses the conceptual characteristics of the cultural and creative tourism industry, focusing on the analysis of the development advantages and opportunities of the cultural and creative tourism industry, the international reference of the cultural and creative tourism industry, the existing scientific and technological achievements and the potential mechanism of the integration of the cultural and creative industry and the tourism industry. Combining with Jiangxi's local cultural resources and tourism resources, carry out creative upgrades, expand product upgrades and market demand, accelerate the transformation of economic development structure, and establish cultural and creative tourism areas that conform to the status quo of Kunming. At the same time, integrating the characteristics of the tourism industry into the cultural creative industry can not only increase the propaganda of the cultural creative industry, increase product sales, and create new forms of cultural creativity and tourism industry derivatives. Play a role in increasing the quality of consumer consumption and promoting the development of integrated industries.

References

1. de León, M. G. U. G., Pereiro, X.: CULTOUR+. An erasmus+ strategic partnership on cultural tourism, entrepreneurship education and innovation. *Int. J. Sci. Manage. Tour.* **3**, 207–242 (2017)
2. Richards, G.: Tourism, an underestimated driving force for the creative economy. *Revista Turismo em Análise* **29**(3), 387–395 (2019)
3. Lopes, E.R., Rego, C.: Cultural and touristic management: cultural and creativity diversity in touristic territories. *Int. J. Sci. Manage. Tour.* **3**, 341–350 (2017)
4. Yamamura, T.: Contents tourism and creative fandom: the formation process of creative fandom and its transnational expansion in a mixed-media age. *J. Tour. Cult. Chang.* **2**, 1–15 (2020)
5. Bae, H.O., Ha, S.Y., Kang, M., Lim, H., Min, C., Yoo, J.: A constrained consensus based optimization algorithm and its application to finance. *Appl. Math. Comput.* **416**, 126726 (2022)
6. Thirugnanasambandam, K., Rajeswari, M., Bhattacharyya, D., Kim, J.: Directed Artificial Bee Colony algorithm with revamped search strategy to solve global numerical optimization problems. *Autom. Softw. Eng.* **29**(1), 13 (2022)

7. Schein, L.: *Tourism and Prosperity in Miao Land: Power and Inequality in Rural Ethnic China*, by Xianghong Feng. Lanham, MD: Lexington Books, 2017. ix+183 pp. US\$90.00 (cloth), US\$85.00 (eBook). *China J.* **79**, 171–173 (2018)
8. Benmamar, B.: Quality of service optimization in orthogonal frequency division multiplexing-based cognitive radio systems based on shuffled frog leaping algorithm. *Concurr. Comput. Pract. Exp.* **34**(1), e6530 (2022)
9. Chahardoli, M., Eraghi, N.O., Nazari, S.: Namib beetle optimization algorithm: a new meta-heuristic method for feature selection and dimension reduction. *Concurr. Comput. Pract. Exp.* **34**(1), e6524 (2022). <https://doi.org/10.1002/cpe.6524>
10. Shukla, A.K.: Chaos teaching learning based algorithm for large-scale global optimization problem and its application. *Concurr. Comput. Pract. Exp.* **34**(1), e6514 (2022)



Construction of an Intelligent Student Management and Evaluation Information System Based on Fuzzy Neural Network

Cheng Qian^(✉)

Student Affairs Office of Party Committee, Wuhan Business University, Wuhan, China

Abstract. With the increasing application of information management systems in colleges and universities. However, specialized information evaluation software for student management is relatively rare, and it is difficult to conduct effective specialized performance evaluation of student management in a traditional information system environment. To this end, the establishment of a complete system for collecting and processing student management information in higher vocational colleges, as well as an information management system built on this basis, has important practical significance for the performance evaluation of student management. The purpose of this article is to study the construction of an intelligent student management and evaluation information system based on fuzzy neural network. In this article, the fuzzy neural network model will be combined to provide new research tools, ideas and methods for the intelligent student management evaluation information system. The intelligent information evaluation system and information management system developed by it also provide new ideas and new methods for colleges and universities. It can also fully mobilize the work of colleges and universities to provide a new platform that conforms to the trend of the times. Experimental research shows that this study uses the fuzzy neural network intelligent student management evaluation information system to evaluate the student information of 6 colleges/departments. The evaluation score calculated only by the weight of the indicator system is more accurate and more flexible.

Keywords: Fuzzy neural network · Student management · Information evaluation · Evaluation system

1 Introduction

In recent years, China's informatization construction has advanced by leaps and bounds, and information systems have been widely used in colleges and universities. As colleges and universities rely more and more on the student information evaluation system, the weakness of the information system is becoming more and more obvious, and the frequency of system interruption caused by the paralysis of the information system is also obvious [1, 2]. At present, the intelligent student management evaluation information can be well carried out through the fuzzy neural network. The combination of the

student management evaluation information system and the fuzzy neural network is of great significance to the development of college education.

In the research on the construction of an intelligent student management evaluation information system based on fuzzy neural networks, many scholars have studied it. For example, Liu RF decomposes the development model of student management evaluation information system into needs analysis, development model establishment, detailed design and function realization Three parts [3]. Zhu Y describes the overall structure of the operation of the student management and evaluation information system, and shows the connection and communication methods of each part, thus showing the realization of the system functions [4]. It can be seen that the research on intelligent student management and evaluation information system based on fuzzy neural network is of great significance.

This paper proposes a fuzzy neural network based on genetic algorithm, which uses the comparative and global search capabilities of genetic algorithm to improve the design of student control and evaluation information system. This article focuses on the system analysis, operation requirements, data flow, data processing and system management evaluation of student management information, and analyzes the main requirements of the program, business relationships and major business planning.

2 Research on the Design of Intelligent Student Management and Evaluation Information System Based on Fuzzy Neural Network

2.1 Fuzzy Neural Network Intelligent Information Processing Combined with Student Management and Evaluation Information System

(1) Principal element network reduction

An important issue in the process of intelligent information processing is to select features with strong distinguishing ability (or the input of training vectors). Therefore, it is necessary to determine the priority of input vector features and extract important features. And use the selected features as input for recognition or classification schemes. At the same time, similar inputs may belong to the same category. Under this rule, the input variables are normalized in the unit interval, and then selected according to their variance. That is, the larger the variance, the more likely the input vector is to have better distinguishing ability [5, 6]. For some data sets, combining two features will produce better distinguishing ability than using one feature alone. Mention that principal component analysis (PCA) (or Karhunen-Loeue transformation) is a method of linearly combining inputs and identifying their priority.

(2) Reduction based on the weight matrix

In the process of data analysis, processing and reasoning, even after the formation of reasoning rules, it is impossible to determine which data or conditional attributes are important or necessary. From the example in the second part of this article, according to the numerical characteristics of the weight matrix, only its second

input attribute has an impact on the perceptron. It can be seen that in the multi-attribute decision-making system, some redundant conditional attributes can be reduced by the weight of the network.

(3) Neural network based on rough reduction

The input information generated by the rough and complex process, in short, may be uncertain quantitative data or powerful phrases, and explain the relationship between site characteristics and decision-making characteristics, that is, the relationship between input points and production sites. To determine the flexibility of the flexible table, rough set reduction can simplify the training data set of the neural network, eliminate redundant data on the premise of retaining important information, and reduce the dimension of data input and the amount of calculation of the neural network.

Currently, there are two types of simplification of neural networks using rough reduction.

1) Use rough reduction to simplify the input dimension of the neural network

The structure of the network is greatly affected by the size of the data input to the network [7, 8]. Therefore, delay reduction is often used to reduce the data size as the input to the neural network system, which can better define the network system.

2) Use rough reduction to simplify the number of fuzzy neural network rules

Fuzzy neural network is a system based on fuzzy inference, which is greatly affected by the number of fuzzy rules. In practical applications, many rules will lead to system integration failure, long network training time or large training errors. Therefore, the use of reducing the number of neural network decision rules can significantly reduce the scale of network configuration and speed up network integration.

2.2 Demand Analysis and Development Mode Selection

The design also has the function of sharing student information to facilitate the openness and transparency of student information, especially attendance, rewards and punishments. It is very important for students to control the personal data and statistical results generated in campus activities. This is based on the implementation of many student departments in campus life. Students can use the IC card to enter the query terminal to fully understand the information they want.

(1) Choice of development mode

The student information intelligent management system based on IC card and network monitoring should use the school's existing campus network and office automation OA system, and add the query terminal and database designed in this project to change the existing system to a lower degree, even It runs in the original system in a parasitic way. The hardware investment is a computer host and a steel case, and this design is mainly composed of software, leaving enough space for function expansion, which can further cover the functions to various application sites and fields in the school [9, 10]. Therefore, the development mode should

be based on general technology, and the operation interface familiar to the public should be selected. To build this design system, you need to make five choices: operating system, background database, scripting language for developing dynamic pages, development mode and system operating environment. Based on the needs analysis and the above considerations, appropriate choices can be made according to the characteristics of various development tools and development languages.

(2) Choice of development structure

The overall design idea of this project is to combine the existing campus network and data system of Guangzhou Public Utilities Senior Technical School to establish a complete student database, log in on the terminal through the student's IC card, and select the desired data on the terminal through touch screen technology. The content to be queried, or the teacher logs in on the office automation system, selects the content to be processed through the browser, and the client selects the collected IC card number or student number to identify the student's identity in the database to filter out the student's designated relevant information Data, and return to the terminal or browser for reference or modification [11, 12].

2.3 Research on Intelligent Algorithm Design

In the evolution strategy, it mainly includes two elements (X, σ). Where X represents the next visited node and σ represents the variance. Its node descendant generation formula is as follows:

$$\sigma' = \sigma eN(0, \Delta\sigma) \quad (1)$$

$$x' = x + N(0, \Delta\sigma') \quad (2)$$

Among them, the mean of $N(0, \Delta\sigma')$ is 0, the variance is σ , and the two are independent and Gaussian random number vectors.

3 Experimental Investigation and Research of Intelligent Student Management and Evaluation Information System Based on Fuzzy Neural Network

3.1 Sources of Experimental Data

Because intelligent evaluation requires experts to score the scientific level of student management evaluation, it is used as training data to train the evaluation model. In terms of student information management in this city, six experts will be selected to sort out the important details of the evaluation information this semester. Based on the basic student management information submitted by each department in January 2020, develop a simulation test data system, train a fuzzy neural network model, summarize the current forecast value of the fuzzy neural network, and evaluate the weight of the system indicators.

3.2 Experimental Method

Due to the reliability characteristics of the index data in the program, 6 experts with rich academic experience and student management experience were selected to evaluate the object and reliability characteristics of the index data in the classroom. In order to evaluate the reliability of the theme and each indicator, subtract the highest score and the lowest score, and get the average value of other experts' statistical data as the final result.

4 Investigation and Analysis of Intelligent Student Management and Evaluation Information System Based on Fuzzy Neural Network

4.1 Analysis of Student Assessment Information

Table 1 lists the student valuations of the five colleges. Table 1 shows the number of times the experts gave the highest evaluation, the lowest evaluation, and the effective evaluation.

Table 1. Student assessment information

Serial number	Highest evaluation	Minimum assessment	Effective evaluation
1	32	26	141
2	20	17	156
3	16	14	178
4	26	14	156
5	18	11	153

As shown in Fig. 1, the system can effectively manage multiple evaluation data and calculate the evaluation scores based on the density of the index system entered into the system by accessing the student information data of 5 faculties and departments through the test. Considering that the evaluation score is only obtained by simple weighting, in order to get a better and reasonable evaluation score, the intelligent evaluation module is further tested. Through this test, the score calculated only according to the weight of the indicator system can be more Accurate and flexible results.

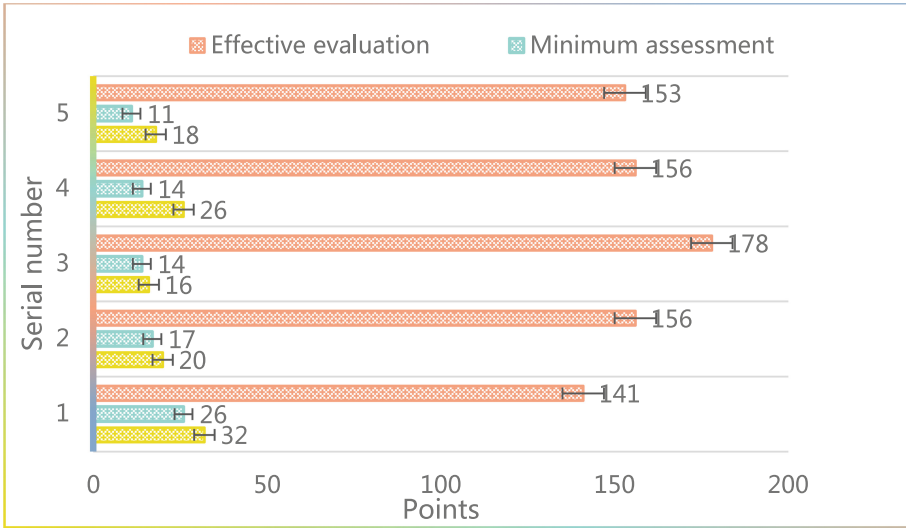


Fig. 1. Student assessment information

4.2 Intelligent Evaluation and Analysis

Since there is no further evaluation data record, the test adopts the (LOO) method. Specifically, the school has 6 colleges and 20 test data. Table 2 records the complete difference between the scores of the two models in the LOO test and the weighted scores of the reference system.

Table 2. Comparison of intelligent evaluation scores

Academy	M1	M2	Weighted score
Economics and Management	81	83	85
International Cooperation and Exchange	84	91	93
Engineering and Information	74	82	85
Humanities and Social Sciences	91	89	82
Tourism	83	79	76
Art and Design	81	75	77

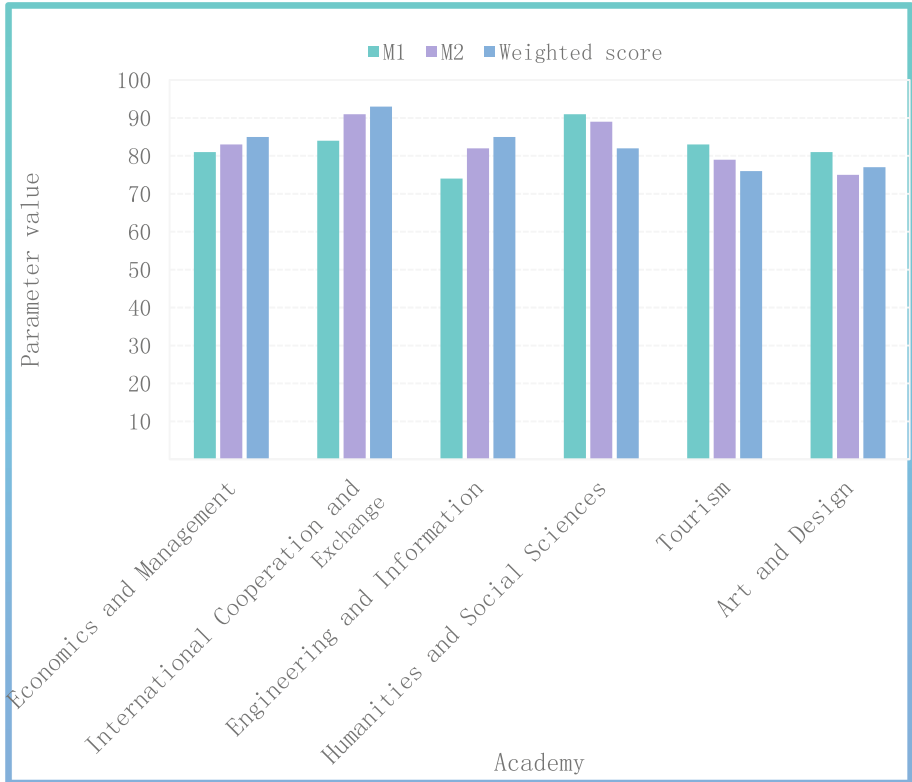


Fig. 2. Comparison of intelligent evaluation scores

As shown in Fig. 2, it can be seen that the results of the model are different from the weight scores of the indicator system. Since the model aims to maximize the experience and personal interests of experts, this difference also shows that the ability of students in different fields to manage evaluation experts is different from the weights in the index system. Utilizing this feeling through the neural network model helps to design a more comprehensive indexing system, and can also be used as an additional benefit of existing indexing systems.

5 Conclusions

For the school's existing intelligent student management and evaluation information system, the teaching department has conducted investigations and studies. During the investigation and research, the educational administrators and the teaching directors in charge of teaching in each college put forward different user needs. According to these users Demand, conduct demand analysis and quantify demand. In the design of the intelligent student management system and the evaluation information system, the functional modules in the intelligent student management system and the evaluation information system are designed respectively. First, it includes the design of iterative

system, the design of cognitive relationship, the system design of test document module and teaching control unit. The system is developed and designed through Java, and an intelligent student management information evaluation system is realized.

References

1. Studiyaniti, S., Azmi, S.: Usability evaluation and design of student information system prototype to increase student's satisfaction (case study: X University). *Ind. Eng. Manage. Syst.* **18**(4), 676–684 (2019)
2. Yin, X.H.: Construction of student information management system based on data mining and clustering algorithm. *Complexity* **2021**(2), 1–11 (2021)
3. Liu, R.F., Wu, M.: Forecast and evaluation of educational economic contribution based on fuzzy neural network. *Complexity* **2021**(2), 1–11 (2021)
4. Zhu, Y., Wang, J.Y., Yang, Y., et al.: Fuzzy evaluation of urban traffic congestion based on neural network. *Beijing Ligong Daxue Xuebao/Trans. Beijing Inst. Technol.* **38**(5), 487–492 (2018)
5. Maekawa, A., Yamakami, M., Kitahara, S.: Analysis of fuzzy neural network control applied to the lateral maneuvering of aircraft. *Nihon Kōkū Uchū Gakkai ronbunshū = J. Jpn. Soc. Aeronaut. Space Sci.* **65**(4), 137–142 (2017)
6. Fuada, S.: Incident management of information technology in the Indonesia higher education based on COBIT framework: a review. *EAI Endorsed Trans. Energy Web* **19**(22), 156387 (2018)
7. Wei, Y., Fang, F.: The design and realization of the management system of college physical education under the network environment. *J. Phys. Conf. Ser.* **1345**(5), 052034 (2019). (5pp)
8. Ozerova, G.P.: Usage of learning management system web analytics in blended learning self-study evaluation. *Vysshee Obrazovanie v Rossii = High. Educ. Russia* **29**(8–9), 117–126 (2020)
9. Purnamasari, F., Ashaari, N.S.: User centered design approach to redesign graduate student management information system. *Data Sci. J. Comput. Appl. Inform.* **2**(2), 109–114 (2018)
10. Haddad, F.S.: Examining the effect of learning management system quality and perceived usefulness on student's satisfaction. *J. Theor. Appl. Inf. Technol.* **96**(23), 8034–8044 (2018)
11. Yeo, S.: Course quality management based on monitoring by students at a medical school. *Korean J. Med. Educ.* **30**(2), 141–152 (2018)
12. Amy, M., Shirrell, M.F.A.: Collaborative learning of diabetes with graphic design communication students and student pharmacists. *Collab. Healthc. Interprof. Pract. Educ. Eval. (JCIPE)* **10**(1), 4 (2019)



Forecasting Model of Number of Entrepreneurs in Colleges and Universities Based on PSO Algorithm

Pingping Yang^(✉)

Nantong Institute of Technology, Nantong 226000, Jiangsu, China

Abstract. The number of college graduates in China has hit a record high, and it coincides with the transition period of the country's economic structure. The employment situation of college students is extremely severe, and the problem of college graduates' entrepreneurship has once again become the focus of society. The purpose of this article is to study the prediction model of the number of entrepreneurs in colleges and universities based on the PSO algorithm. Introduced the research background and significance of data prediction, summarized the research status of data prediction at home and abroad and the research status of BP neural network. Faced with the problems in the process of BP optimization by particle swarm optimization, an improved PSO-BP prediction model is proposed. This model uses BP to update the optimal position every time the particle swarm is iterated. The new combination prediction model is used in the college student entrepreneurship prediction experiment. The simulation results show that the predicted result is that the number of construction college students starting a business in 2022 is 880,000, and the number of construction college students starting a business in 2023 is 1.07 million. The combined forecasting model has better forecasting effect.

Keywords: PSO algorithm · Number of entrepreneurs · Prediction model · Number of universities

1 Introduction

Entrepreneurship has grown in recent years because it has been fully embodied in economic growth, employment promotion, and innovative value [1]. In fact, the expansion of college enrollment in recent years has brought difficulties to employment, so embarking on the road of independent entrepreneurship is a new value and career direction for current college students [2]. In view of the external environment of society, it is inevitable to accelerate business research. Cultivating more entrepreneurial talents and promoting the development of the business economy have important practical significance for the government and education departments [3, 4].

Particle swarm algorithm has fast convergence speed and simple operation [5]. Liu W gave the meteorological model required by the SVM model. Then use the selected mode consisting of relative humidity, wind speed and wind direction to obtain higher

prediction accuracy. The (PSO) algorithms are studied to optimize the parameters in the SVM classification process [6]. Tian Z proposed an improved particle swarm optimization algorithm to reduce the number of useless neurons to achieve the purpose of reducing network complexity and improving algorithm efficiency. The results shows the prediction model proposed in this paper reduces the training time [7]. At present, most of the domestic researches on college students' entrepreneurship are focused on college students' entrepreneurship education, and there are few studies on the number of college students' entrepreneurship predictions.

The innovation of this article: It is difficult for people to find a definite formula to predict complex and changeable entrepreneurial activities. In the field of entrepreneurship, there are relatively few studies on entrepreneurial intentions. Moreover, few domestic scholars have conducted in-depth research on the influence of emotional prediction bias on individuals. Even foreign scholars have hardly introduced this important concept into the field of entrepreneurship. Therefore, this paper proposes a network traffic prediction model based on PSO-BP algorithm for the shortcomings of BP neural network. And through this forecasting model, a network traffic forecasting management system is designed and implemented, which can realize functions such as traffic forecasting and strengthen the effectiveness and speed of network management.

2 Research on the Forecast Model of the Number of Entrepreneurs in Colleges and Universities Based on PSO Algorithm

2.1 Entrepreneurship in Colleges and Universities

Entrepreneurship can be divided into broad sense entrepreneurship and narrow sense entrepreneurship. Entrepreneurship in the broadest sense refers to people's creative activities or social activities that have positive significance for pioneering and innovative activities. This type of activity can be for-profit or non-profit [8, 9]. It can be economic, political or cultural. People can choose their own positions according to their interests and hobbies, and realize their personal value by fully demonstrating their skills and talents at work. Since they are all entrepreneurs, careers are also very important for entrepreneurship. Traditionally, we tend to study entrepreneurship from a narrow perspective. Entrepreneurship in universities means that students do not seek employment through traditional employment channels after graduation, but rely on their own knowledge, wisdom, patent achievements and other resources, independently or in other ways [10, 11].

Entrepreneurship in universities has the same essence as other groups. As a member of an enterprise team, students must insist that ordinary entrepreneurs must integrate into the core characteristics of the enterprise group. Based on the opinions of related scholars, this research integrates entrepreneurial optimization into a group of resources that students can have through their own resources or hard work, and defines the process of increasing economic or social value as a production process [12].

There are many studies on the concept of entrepreneurship, but they believe that the basic elements of entrepreneurship are basically the same. This is a personal belief, commitment and plan for entrepreneurship. This research combines many concepts

of entrepreneurial intention in academia and finds that entrepreneurial intention is the degree of potential entrepreneurial psychological preparation for entrepreneurial activities or business behaviors.

2.2 Combination Forecasting Model of the Number of Entrepreneurs in Colleges and Universities Based on PSO Algorithm

Using particle swarm algorithm to optimize BP neural network has the advantages of fast convergence, simple operation, high efficiency, and strong generalization ability. When Gbest is the local optimal, the optimal Pbest position will cause the deviation of the particle direction. And speed, as a result, you cannot get the overall optimal solution. The standard PSO optimization BP process is prone to partial optimization, which will lead to faster convergence and less accurate predictions. The optimal position updates the optimal positions of the Pbest and Gbest particle swarms. The operation process is as follows:

- (1) Initialize the neural network and particle swarm. First of all, it is necessary to determine the structure of the neural network and the algorithm deployment in the network, and determine the dimension of a single particle according to the number of neural network weight thresholds. Determine the range of the weight threshold, and randomly generate the position sequence and velocity sequence in the particle according to the range.
- (2) Calculate the fitness of each particle according to the error of the neural network, and update the optimal position Pbest of each particle and the overall optimal position Gbest according to the fitness.
- (3) Put the particles in the particle swarm into the neural network one by one, and perform a loop training to obtain new particles. The new particles will update the optimal position of the particle swarm and the optimal position of the particle without affecting the original Some particle swarms.
- (4) After updating the optimal position of the individual and the optimal position of the group, update the position sequence and velocity sequence of the original particles according to the formula. After the update is completed, judge whether the current iteration number reaches the maximum iteration number, if not, execute the second step, otherwise execute the fifth step.
- (5) Use the optimal particle swarm optimization as the neural network weight threshold, create a new neural network, and use the network to predict the test data. In the process of BP neural network optimizing the particle swarm, the movement of the particles in the swarm is better. Program to change speed and position. Instead of blindly moving to the local best solution, but moving in the direction of the best solution. The improved PSO-BP model can reduce the chance of particle droplets falling into the early local optimization, making the BP neural network prediction more accurate and faster convergence.

3 Investigation and Research on Forecasting Model of Entrepreneur Number in Colleges and Universities Based on PSO Algorithm

3.1 Data Sample

Taking the 2021 undergraduates of sample universities as the survey object, the entrepreneurial quality scores of 2021 undergraduates of sample universities are calculated according to the built-up evaluation model of university graduates' entrepreneurial quality. By analyzing the various indicators of employment quality, it is found that there are problems in the entrepreneurship of graduates. The data source of this research is divided into two parts: one is to obtain the basic entrepreneurial situation of 2021 undergraduate graduates through communication with the entrepreneurial guidance center of sample universities, including entrepreneurial guidance, income, and basic information about the startup company. And four influencing factor indicators such as the professional impact factor of learning are used as the input sample P, and the number of college students starting a business in 2021 is used as the target sample T to establish a relevant data set. The relevant data in 2021 will be used as training samples for the BP neural network. When the BP neural network is programmed, the relevant data is normalized using the Premnmx() function that comes with the Matlab software, so that all relevant variables are kept between [0,1].

3.2 Data Preprocessing

The fitness of the new particles is calculated, and then P_{best} and the optimal position of the group G_{best} are updated with the formula:

$$P_{best}(i) = \begin{cases} P_S(i), F_S(i) < F_{best}(i) \\ P_{best}(i), F_S(i) \geq F_{best}(i) \end{cases} \quad (1)$$

$$G_{best} = \begin{cases} P_S(i), F_S(i) < F_{G_{best}}(i) \\ G_{best}, F_S(i) \geq F_{G_{best}}(i) \end{cases} \quad (2)$$

In the formula, I is the particle number, $P_{best}(i)$ is the optimal position of the i particle; $P_S(i)$ is the position of the i particle after the BP change; $F_{best}(i)$ is the fitness of the optimal position of the i particle Value; $F_S(i)$ is the fitness value of the new particle i; $F_{G_{best}}$ is the fitness value of the optimal position of the group.

4 Investigation and Research Analysis of the Predictive Model of the Number of Entrepreneurs in Colleges and Universities Based on the PSO Algorithm

4.1 Predictive Model Verification

Select the data on the number of undergraduate entrepreneurs from 2016 to 2020 and the corresponding data on the 4 factors that affect the change in the number of undergraduate entrepreneurs, and carry out the training of the BP neural network model. After the

training, the actual number of undergraduate entrepreneurs from 2016 to 2020 will be brought back into, Use the simulation function sim function to simulate and see whether the error requirements are met. The actual value and the simulated output value can be obtained through the Matlab software analysis result as shown in Table 1.

Table 1. Comparison results between the actual number of undergraduate entrepreneurs from 2016 to 2020 and the simulated predicted value

Years	Actual value	Analog value	Relative error
2016	1.01 million people	1,002,65 million	0.000826
2017	2.21 million people	2,217,80 million	0.000297
2018	2.99 million people	2,988,60 million	0.002553
2019	3.86 million people	3,850,72 million	0.001162
2020	4.3 million people	4,399,60 million	0.000607

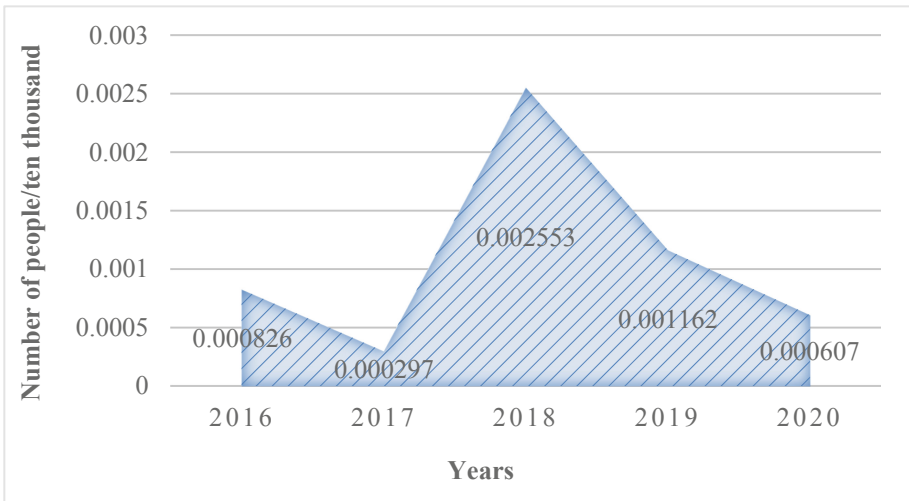


Fig. 1. Relative error

The relative error of this model satisfies the requirements as shown in Fig. 1. The relevant data from 2016 to 2020 is used as the detection sample of the BP neural network. After the training simulation is over, the factors that affect the number of college students' entrepreneurship changes from 2016 to 2020 are brought in, and the simulated predicted

number of college students' entrepreneurship in 2021 is 5.06 million, and the actual number of college students' entrepreneurship in 2021 is 5.05 million. The relative error is 0.0447. From 2016 to 2020, it has reached an accurate fitting state under the condition of training supervision, and the simulated prediction value in 2021 has a certain error, but the error is still controlled within 0.1, indicating the established BP The prediction accuracy of the neural network prediction model is extremely high, and it is suitable for scientific prediction of the number of migrant workers returning home to start a business in 2022 and 2023.

4.2 Prediction Results of the Number of Undergraduate Entrepreneurs

The gray GM(1,1) model is used to scientifically predict the value of various factors that affect the number of college students' entrepreneurship changes in 2022 and 2023, and the proportion of the number of entrepreneurs to all college students. The results of the prediction are shown in Fig. 2.

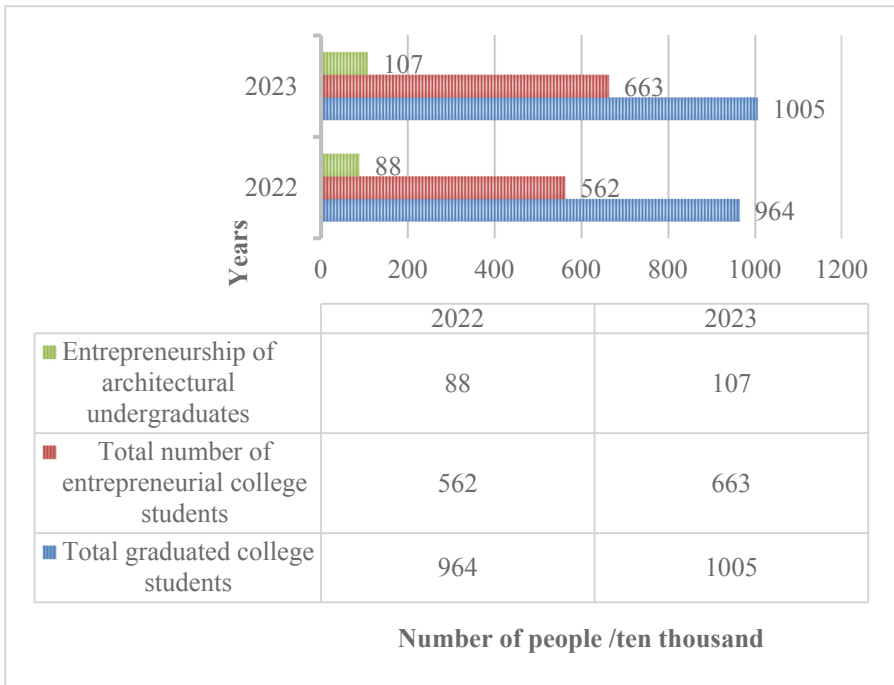


Fig. 2. Gray GM(1,1) model predicted value

The trained BP neural network is used to predict the number of undergraduate entrepreneurs in 2022 and 2023, and the result is that the number of undergraduate entrepreneurs will continue to grow in 2022 and 2023, and the number of undergraduate entrepreneurs in 2022 will be 5.62 million and 6.63 million in 2023.

It is proposed to adopt the proportion of the number of entrepreneurs of architecture majors in the number of entrepreneurs of all college students as the forecast of the number of entrepreneurs of architecture majors in 2022 and 2023. Specifically, the gray GM(1,1) model predicts the average of the proportion of the number of architectural college students in the total number of college students in 2022 and 2023 as the forecast of the number of construction college students in 2022 and 2023, and the predicted result is In 2022, the number of architectural college students starting a business is 880,000, and in 2023, the number of architectural college students starting a business is 1.07 million.

5 Conclusions

In this research, in order to introduce the concept of number prediction in the field of entrepreneurship and improve the accuracy of data prediction, we focus on current artificial intelligence algorithms and study data prediction. This is a unique quantitative prediction model. In this article, we studied the BP neural network and its optimization method, proposed an improved PSO-BP prediction model through comparison, and used the Adaboost algorithm to combine the model with other models to create a new prediction coupling model. Applied empirical research not only focuses on the factors affecting the number of entrepreneurs mentioned in previous studies, but also provides a useful reference for quantitative empirical research on future business intentions, and also adds more realistic and specific explanations. This research may make up for some of the shortcomings of previous research and is of great significance to provide specific theoretical guidance.

References

1. Alon-Beck, A.: The coalition model, a private-public strategic innovation policy model for encouraging entrepreneurship and economic growth in the era of new economic challenges. *Wash. Univ. Glob. Stud. Law Rev.* **17**(2), 5 (2018)
2. Kwilinski, A., Litvin, V., Kamchatova, E., et al.: Information support of the entrepreneurship model complex with the application of cloud technologies. *Int. J. Entrep.* **25**(1), 1–8 (2021)
3. Huang, Z., Liu, G.: Prediction model of college students entrepreneurship ability based on artificial intelligence and fuzzy logic model. *J. Intell. Fuzzy Syst.* **40**(2), 2541–2552 (2021)
4. Shokan, R., Karipova, A., Khoich, A., et al.: Economic levers of regulation of entrepreneurship. *J. Entrep. Sustain. Issues* **7**(2), 1567–1581 (2019)
5. Pinkovetskaia, I.S., Nuretdinova, Y.V., Navasardyan, A.A.: Entrepreneurship in the third age: data from Russia. *Revista Amazonia Investiga* **9**(27), 165–172 (2020)
6. Liu, W., Guo, G., Chen, F., et al.: Meteorological pattern analysis assisted daily PM2.5 grades prediction using SVM optimized by PSO algorithm. *Atmos. Pollut. Res.* **10**(5), 1482–1491 (2019)
7. Tian, Z., Ren, Y., Wang, G.: Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM. *Energy Sources* **41**(1–6), 26–46 (2019)
8. Al-Molegi, A., Jabreel, M., Martinez-Balleste, A.: Move, attend and predict: an attention-based neural model for people's movement prediction. *Pattern Recogn. Lett.* **112**(SEP.1), 34–40 (2018)

9. Zoest, R., Law, M., Sabin, C.A., et al.: Predictive performance of cardiovascular disease risk prediction algorithms in people living with HIV. *JAIDS J. Acquir. Immune Defic. Syndr.* **81**(5), 1 (2019)
10. Nijhawan, A.E., Metsch, L.R., Song, Z., et al.: Clinical and sociobehavioral prediction model of 30-day hospital readmissions among people with HIV and substance use disorder: beyond electronic health record data. *JAIDS J. Acquir. Immune Defic. Syndr.* **80**(3), 1 (2018)
11. Song, G., Min, S., Lee, S., et al.: The effects of network reliance on opportunity recognition: a moderated mediation model of knowledge acquisition and entrepreneurial orientation. *Technol. Forecast. Soc. Change* **117**(APR), 98–107 (2017)
12. Byeon, H.: Development of a physical impairment prediction model for Korean elderly people using synthetic minority over-sampling technique and XGBoost. *Int. J. Adv. Comput. Sci. Appl.* **12**(1), 36–41 (2021)



Design of Automatic Evaluation of Machine English Translation Based on BP Neural Network

Yan Zhang^(✉)

Zaozhuang University, Zaozhuang, Shandong, China

Abstract. In recent years, with the development of deep learning, the automatic evaluation model of neural network machine translation based on deep learning has been widely used. In addition, the task of machine translation quality assessment is closely related to machine translation. The purpose of this article is to study the design of automatic evaluation of machine English translation based on BP neural network. This article proposes a method to extract features containing translation knowledge with the help of a neural network machine translation model. This article evaluates the quality of machine translation from the perspective of language error analysis theory, analyzes the errors and their causes, and helps provide quality support for machine translation research from the perspective of computer science. Experimental studies have shown that as the scale of the data set grows, the accuracy of both BLEU value and n-grams reaches about 99.3, and the model effect is significantly improved.

Keywords: Neural network · Machine translation · Deep learning · Translation quality evaluation

1 Introduction

Among the existing research results of English machine translation engineering, the main research direction is to study the engineering progress of machine translation from the perspective of computer technology. This article is dedicated to providing high-quality and advanced support for machine learning research from the perspective of BP neural network, which plays an important role in automatically judging the neighborhood of machine translation.

In the research on the automatic evaluation design of machine English translation based on BP neural network, many scholars have studied it and achieved good results. Phrase segmentation, etc., use features to describe the rules of translation, and use dynamic algorithms to achieve efficient translation evaluation [1]. Ren Q realizes automatic translation between natural languages through BP neural network, which has become the mainstream translation technology [2]. Therefore, it is very necessary to study the automatic evaluation design of machine English translation based on BP neural network.

The research of this article is divided into three main stages: preparation stage, analysis stage and summary stage. In the preparation phase, the object will obtain the text and insert the text into the neural network translation system to obtain the translation language results: machine translation errors, then quantitative calculations and error intensity analysis, and then try to analyze the reasons for the errors; in the final stage, the translation Quality is assessed.

2 Research on the Design of Automatic Evaluation of Machine English Translation Based on BP Neural Network

2.1 Features of Fusion Translation Knowledge

In the process of feature extraction, the BP neural network machine translation model is used to make full use of a large number of parallel corpus, and the extracted features incorporate translation knowledge. Then, after extracting the above features, it is input into a simple QE model, and finally the prediction result HTER is output. Here, a forward BP neural network with a single hidden layer is selected as the structure of the QE model.

- (1) First, all model parameters are randomly created, including text vector parameters.
- (2) Construct two NMT models. It can be done in two ways, alternative training and joint training. Another training method is as follows, the existing custom NMT_A template and NMT_B open source template. Train multiple levels for NMT_A and push these levels into the queue. After NMT_A repeats multiple levels, start training NMT_B, and repeat the levels in the training queue before exiting NMT_B.

2.2 Transformer Model Design with Part-of-Speech Information Vector

- (1) Principle of Transformer translation model

The first is the encoder part. The biggest feature of the cyclic BP neural network is that it can reflect the positional relationship of the input sequence itself. For natural language processing, the positional relationship of each word in the sentence is already included in the input [3, 4]. In the Transformer model, there is no cyclic BP neural network in the BP neural network it uses, so it cannot reflect the position information between the input words in the sentence, which means that no matter where the input words are input, there will be no difference.

In order to solve this problem, in the Transformer model, researchers will use two methods when converting the source language into vectors—one is the commonly used word embedding, and this step is to convert the word The text representation is mapped to the numerical space. Because the computer only accepts numerical data, this is a necessary step in processing text information; another way is the positional encoding that can reflect the position information, in which the position of the word is passed The distance between the word and other words is reflected.

- (2) Motivation and model design

- 1) The motivation for adding part-of-speech information vector

First of all, in Transformer's original model, two vectors are used in the process of converting text into vectors when inputting the source language. The two vectors are

added to obtain the final input vector, which is a regular word vector representation. This method Map the text representation of the word in the sentence to the numerical space and convert it to the representation of a numerical vector. This mapping is to allow the computer to read the numerical information of the text; the second is the representation of the position information vector. This method is based on the word the position in the sentence where it is located generates a position vector, where the position is determined according to the distance between the word and other words in the sentence.

2) Transformer model design of part-of-speech information vector

This article believes that since machine translation is ultimately a problem of translation, and the understanding of part-of-speech information is very important for translation problems, the effect of the model after adding part-of-speech information will be improved [5, 6]. The first input word text will go through the step of word embedding (word embedding is to convert the language from the expression of the text to the expression of the value, in fact, it is a kind of mapping, a certain word in the text space is mapped by a certain method Numerical vector space), and then encode the position information and part-of-speech information of the input word. After the three are added, it enters the encoder. The output of the encoder enters the decoder. At the same time, the current output of the decoder will be used as the next step. Input, and will be processed the same as the input of the encoder.

2.3 Research on Neural Machine Translation Model

(1) Building the encoder-decoder framework

The neural machine translation used in this article uses an end-to-end encoder-decoder framework to instantaneously model the translation probability, process the source language sentences as a whole, let the machine automatically learn features, and use the BP neural network for natural language applications. The direct mapping between the methods translates the translation problem into how to determine the conditional probability of the creation of this transfer target language [7, 8].

(2) Neural machine translation model

The memory module, input gateway, receiving gateway, and output gateway have been added to LSTM. The memory module is responsible for storing content and memory on the network: the gateway controls the flow of information to the memory module: the gateway selects historical information to be discarded in the memory module. The output port specifies the current input information and the content of the current memory module output information. The combination of these ports and memory modules significantly improves the ability of repeated BP neural networks to process large data sequences.

(3) Translation model integrating attention mechanism

With the advancement of translation, the part of attention will also change, but the use of a fixed-dimensional vector here means that the same degree of attention is applied to each part of the source language sequence, which is obviously not conducive to the improvement of machine translation performance. In order to solve the above problems, the attention mechanism is integrated into the decoder

to improve the feature learning ability of long sentences [9]. The decoder model with the attention mechanism strengthens the representation ability of the source language sequence. The neural machine translation model is divided into three parts: the encoder, the decoder and the attention mechanism. A set of multiple vectors is used to replace the fixed vector to represent the source. Language information, the background vector is dynamically selected during the target sequence generation process, and the decoding process pays more attention to the part of the source language that is more relevant to the target sequence.

(4) Translation model based on two-way GRU

When performing translation tasks, the context information of the source language translation at the current moment should be considered comprehensively. The cyclic BP neural network can store timing information by sequentially scanning the input sequence, and has the ability to remember context information. However, the word vector obtained by this method only contains the context information on the left side, and does not consider the context information on the right side.

(5) Evaluation criteria for machine translation-BLEU score

The effect of automatic translation can be judged by manual evaluation, but in order to save human resources, the translation quality is calculated by comparing the ratio between the interpretation constructed by the machine model and the interpretation given to the original comparator [10]. The effect of machine learning is usually evaluated by the BLEU score, which is the most commonly used standard for evaluating scores.

2.4 Design of Parallel Classification Model of Bp Neural Network

Based on the training rules of the BP neural network, in the forward propagation stage of the working signal, for the i -th neuron in the k -th layer, the input value is:

$$X_i^k = \sum_{j=1}^{n+1} W_{ij} X_j^{k-1} \quad (1)$$

The output value of this neuron is:

$$Y_i^k = f(X_i^k) \quad (2)$$

In the reverse transmission phase of the error signal, the value range of the expected output value Y_i^k is defined as $[1, N]$.

2.5 Research on Machine Translation Errors

The error correction suggested in this document is a debugging program for machine translation, which ultimately improves the translation quality of automatic translation. Therefore, “misunderstanding errors” is accepted as one of the key points of error classification, and the names of errors with important information are shared. It is a “translation error” or “missing information error”, not a “word error”.

(1) Main body error

From the statistics of the frequency of such errors, the frequency of ontology errors is calculated for a small part of the total number of errors, but the recognition errors of the original text and sentence parts are very concentrated, which is much higher than other types of recognition errors. In terms of machine translation, Google Translate performed very well, surpassing the other two in terms of original word recognition and sentence sharing errors. We can see that although Google Translate has made some achievements in Chinese identity, the issue of Chinese identity and keyword sharing is still the focus of search engines.

When recognizing and splitting sentences, the main error of automatic translation is that the accuracy of Chinese phrases cannot be determined. It is essentially a parallel system. Ontological errors may not indicate any effective English translation from a literal perspective. Incorrect sentences can lead to lengthy machine translation and make it difficult for readers to read.

(2) Allegorical errors in interpretation errors

Definition errors are divided into four categories: word errors, vocabulary misuse, unclear sentence scope and stylistic misuse. From the statistics of the frequency of various types of errors, text errors are at the top of the list, but in the field of automatic translation, text problems are not significant. Vocabulary is a limitation of language. The choice of words is very important to the quality of meaning, and it is the basic basis for the formation of conjunctive sentences.

From the general explanation of the content, the error problem is one of them. It is understandable that due to the tremendous progress of artificial intelligence technology, the translation system is still working hard to use and improve the end-to-end network interpretation model, but it is not urgently needed for success. The Chinese original text has made many texts before the technology has arrived. Make guesses and adjustments outside of the text, and try to avoid changing the subject when the original subject is correctly identified.

(3) Part of the error message

In English, the same root, different roots and terms can be converted into different parts of speech. For example, verbs, nouns or adjectives, the word “flexibility” has many meanings. This kind of Chinese and English vocabulary is rarely fully compatible, manual translation can be easily adjusted and adapted, while automatic translation relies on a more advanced foundation to deal with. Some word errors mainly occur in inconsistencies between words, phrases, adjectives, etc.

3 Research on Experimental Design of Automatic Evaluation of Machine English Translation Based on BP Neural Network

3.1 Experiment Preparation

In order to compare the model effect of word and character level input and verify the influence of the data set size on the model effect, in the experimental stage, two data sets were extracted. The small-scale data set has a magnitude of 3 million, including, the large-scale data set contains 21 million Chinese-English bilingual training sets, and

the test data scale is 10,000. Due to the conversion of the input form from words to characters, the sequence length has increased exponentially. In order to prove that the sequence length has a significant impact on the translation effect, the sentence lengths at the word level and the character level in the large and small data sets are respectively counted.

3.2 Experimental Method

Corresponding to the different calculation results of the BLEU evaluation method. The first item, BLEU, is the main evaluation index. BP is determined by the length, and P1 to P4 represent the accuracy rates of unary words to quaternary words, respectively. In addition, the char1 model corresponds to the evaluation results of the million-level data set, and the char2 corresponds to the tens of millions of level. The model effect is significantly improved when the scale of the data set increases.

4 Experimental Analysis of Automatic Evaluation of Machine English Translation Based on BP Neural Network

4.1 Classification and Statistical Analysis of Sentence Length

The input changes from words to characters, and the input length has doubled. In order to prove the impact of sentence length on the classification effect, this paper classifies the model according to sentence length to test the score. The results are shown in Table 1.

Table 1. Results classified by sentence length

Test set sentence length (characters)	BLEU	P1	P2	P3	P4
1 to 9	20.54	42.64	27.73	16.64	11.93
10 to 19	25.67	48.01	33.36	19.92	13.48
20–29	24.89	43.88	26.79	19.03	12.69
30–39	16.63	33.95	18.81	12.14	9.77
40–49	11.23	27.89	15.51	9.93	4.43
50–59	6.78	24.56	11.47	4.39	1.53

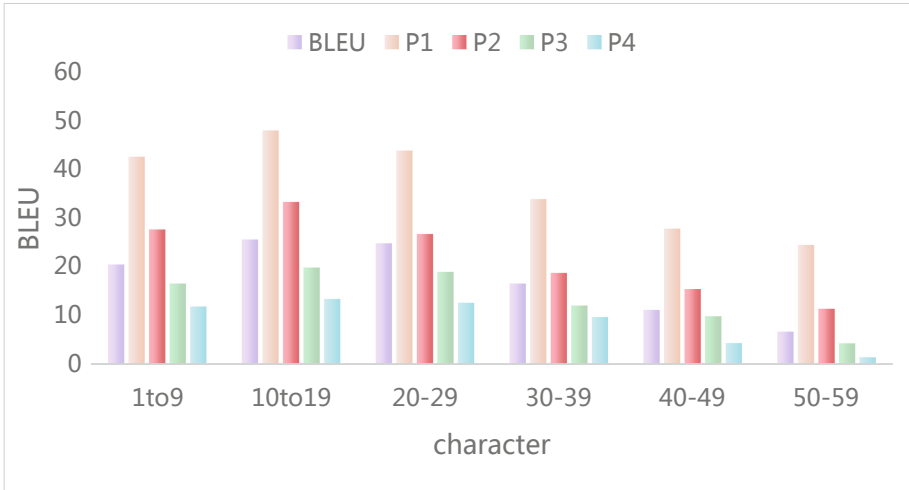


Fig. 1. Results are classified by sentence length

As shown in Fig. 1, when the length is less than 20, the score is higher. As the length of the sentence increases, the score gradually decreases. Therefore, it is necessary to further strengthen the processing ability of long sentences.

4.2 Comparative Analysis of Model Results of Different Data Scales

In order to further analyze and verify the impact of the growth of the data set on the model effect, we used 2 million, 8 million, 14 million, and 21 million data sets as the data sets for training, and evaluated the model. The specific results are shown in Table 2.

Table 2. Comparison of results of different data scale models

Data set size (pair)	BLEU	P1	P2	P3	P4
3 million	19.63	38.03	22.77	14.35	9.99
9 million	20.07	40.96	25.34	15.54	10.47
15 million	20.78	40.35	24.91	16.88	10.15
21 million	22.46	42.07	28.13	18.97	21.31 s

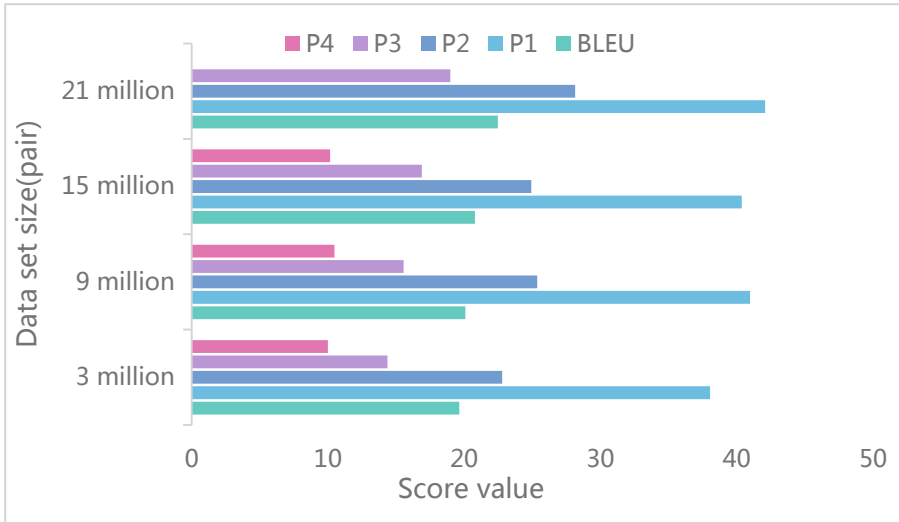


Fig. 2. Comparison of results of different data scale models

As shown in Fig. 2, as the scale of the data set grows, the accuracy rate of both the BLEU value and the n-gram word reaches about 99.3, and the model effect is significantly improved.

5 Conclusions

Compared with statistical machine translation, neural machine translation has a significant increase in complexity. The conversion from characters to words doubles the length of the input, which further increases the difficulty of model training. How to optimize the model structure so that it has less training complexity to improve the processing effect of long sentences is the work that needs to be completed in the follow-up.

References

1. Tezcan, A., Hoste, V., Macken, L.: A neural network architecture for detecting grammatical errors in statistical machine translation. *Prague Bull. Math. Linguist.* **108**(1), 133–145 (2017)
2. Ren, Q., Su, Y., Wu, N.: Research on Mongolian-Chinese machine translation based on the end-to-end neural network. *Int. J. Wavelets Multiresolut. Inf. Process.* **18**(01), 46–59 (2020)
3. Nagaraj, P.K., Ravikumar, K.S., Kasyap, M.S., et al.: Kannada to English machine translation using deep neural network. *Ingénierie des Systèmes D Information* **26**(1), 123–127 (2021)
4. Wang, N., Wang, M.: Research on machine translation sentence pattern transformation method based on BP neural network. *Revista de la Facultad de Ingenieria* **32**(3), 693–700 (2017)
5. Sarker, G., Ghosh, S.: A convolution neural network for optical character recognition and subsequent machine translation. *Int. J. Comput. Appl.* **182**(30), 23–27 (2018)
6. Wu, S., Zhang, D., Zhang, Z., et al.: Dependency-to-dependency neural machine translation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **26**(11), 2132–2141 (2018)

7. Xu, Z., Qin, H., Hua, Y.: Research on Uyghur-Chinese neural machine translation based on the transformer at multistrategy segmentation granularity. *Mob. Inf. Syst.* **2021**(3), 1–7 (2021)
8. Casas, N., Costa-Jussà, M.R., Fonollosa, J., et al.: Linguistic knowledge-based vocabularies for Neural Machine Translation. *Nat. Lang. Eng.* **27**(4), 1–22 (2020)
9. Kim, H., Jung, H.Y., Kwon, H., et al.: Predictor-estimator: neural quality estimation based on target word prediction for machine translation. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **17**(1), 1–22 (2017)
10. Tan, Z., Su, J., Wang, B., et al.: Lattice-to-sequence attentional Neural Machine Translation models. *Neurocomputing* **284**(APR.5), 138–147 (2018)



Design and Implementation of Multimedia Vocal Music Learning System Based on Network Information Technology

Fang Hui^(✉)

Zaozhuang University, Zaozhuang, Shandong, China

Abstract. With the continuous development of network information technology, the society and economy of traditional education need new technology to meet more and more challenges. How to use modern multimedia to develop vocal music education is becoming more and more important. The development of multimedia and educational networks has attracted great attention. This paper aims to study the design and implementation of multimedia vocal music learning system based on network information technology, hoping that this educational model will inject new vitality into educational reform. This paper firstly introduces the theory of network information technology P2P and multimedia system, in addition, as well as multimedia vocal music learning system module requirements and system framework are designed, finally design a multimedia vocal music learning system. The experimental results show that the construction of multimedia vocal music learning system makes the education more convenient and efficient, among which the highest reaches 88%.

Keywords: Network information technology · Multimedia technology · Vocal music learning system design

1 Introduction

With the rapid development of information technology, starting from the basic needs of vocal music teaching, many multimedia technologies can be obtained [1], such as sound wave graphic display, sound spectrum analysis technology and many other aspects [2–5]. Vocal music teaching is an important part of music education in colleges and universities. Teaching vocal music can guide students' vocal music, singing skills and aesthetic appreciation. Moreover, teaching vocal music can also exercise students' thinking in different ways and greatly improve their artistic achievements and creativity [6].

With the rapid development of information technology, traditional basic teaching tasks have not met the high demand of modern music style and interactive education. However, the development of computer science, network technology, multimedia and unified technology has undoubtedly provided a perfect solution for the modernization of education [7]. However, in Hu Q's research, music teachers have sufficient theoretical knowledge and practical experience in music, but their understanding of computer is

relatively rare, and they cannot systematically master computer multimedia technology [8]. Secondly, the multimedia technology of most platforms is decentralized and single use of their own software and hardware systems. Multimedia technology can only be used to analyze the voice of specific departments or multiple departments, but cannot provide systematic voice teaching services [9]. In view of the above issues, the aim is to stimulate future research on the topic. His research direction is very suitable for the current needs, but it is obviously insufficient in terms of practicality.

The innovation of this paper lies in the design and research of an interactive multimedia audio music teaching system, which is an urgent problem to be solved in music teaching. This paper systematically designs and develops the actual pronunciation of multimedia music teaching, which also makes students more interested in vocal music learning and makes the teaching mode of teachers easier to operate [10].

2 Multimedia Vocal Music Learning System Based on Network Technology

2.1 Multimedia System

Complete classroom multimedia equipment includes computers, headlights, electronic curtains, central control system, sound system, microphone, sound, physics room and other equipment. Before multimedia equipment, most college classrooms were mainly controlled by local media, control panel for each program of control device or personal control, working principle is relatively simple and easier to implement. An intelligent multimedia system is a network of intelligent control systems installed in each multimedia room. All relevant multimedia central control devices are installed through real-time detection sensor devices to obtain and analyze the operating conditions of each device. At the same time, status information is transmitted to the multimedia system server database. So many multimedia classrooms are convenient and efficient in use and maintenance. In the Internet technology these intelligent terminals become important items can be easily found and automatic analysis, information exchange, resource management, on-site monitoring, remote control. This is done by using basic Internet technologies to collect information from all data anywhere connected to the network, and the collected information is transmitted over the network to central information processing, information intelligence analysis and cloud computers, and information feedback.

2.2 Peer-to-Peer (P2P) Network

P2P network resources and services are allocated to all network nodes, information transmission and service implementation directly between nodes, no intermediate links, no server intervention, to avoid possible network congestion points. Decompress its best potential in network scalability, robustness and so on. As for P2P networks, each node is a server and a client. By distributing resources among multiple nodes, capacity requirements such as server computing, storage, etc. on traditional C/S nodes are reduced and security and other loads are improved. For P2P network users, you can search for any active node on the network and download directly from that node. Network users can

also share information downloaded from their devices for use by other users. In this way, information about higher demand rates can be quickly distributed across multiple nodes. As the size of the network increases, the amount and scope of shared information will increase. In an open network environment, P2P networks can collect a lot of information very early on. The Internet node to point and does not reflect the content to be presented in real time, whereas on the network, the point information stored at a specific address is transparent to the user, the user is requesting the network, and the application software will call the node's address deposit to translate this information. The user requests an information identifier rather than a physical address. Because information is equivalent to multiple nodes storing this information, it is easier to find information resources by dividing the information distribution according to the content on the network. Table 1 compares the performance of P2P technology and C/S structure from several aspects such as administrative security.

Table 1. P2P and C/S performance comparison

	P2P	C/S
Manageability	Bad	Good
Security	Bad	Good
Fault tolerance	Good	Bad
Scalability	Good	Bad

2.3 Multimedia Network Vocal Music Teaching System

Multimedia network vocal music teaching system is a software system which provides comprehensive teaching service based on interactive multimedia communication network. Through the multimedia and network teaching system, students not only learn the course content, but also carry out a variety of teaching activities, such as communication between teachers and students, communication between teachers and students, submitting homework, answering questions and testing. The system adopts P2P network center standard, the server provides the communication information forwarding service for the node, and adopts the communication mode, the server forwarding adopts multi-thread technology, which is used to deal with the service requests issued by many users at the same time. Vocal music learning teaching subsystem provides a series of online teaching support tools, such as learning content introduction, evaluation function, online question and answer, online examination and so on. Secondly, it also includes resource management system, education management system, curriculum innovation and development system. The educational management subsystem mainly includes teacher management, student management, curriculum management, authority certification and other educational functions. Curriculum development subsystem implements curriculum development and realizes the management and reuse of teaching resources. The resource management subsystem takes the discipline as the unit, and centrally manages all kinds

of teaching resources, including multimedia hardware, test questions, typical cases, bibliography materials, etc. Its multimedia network vocal music teaching system structure is shown in Fig. 1:

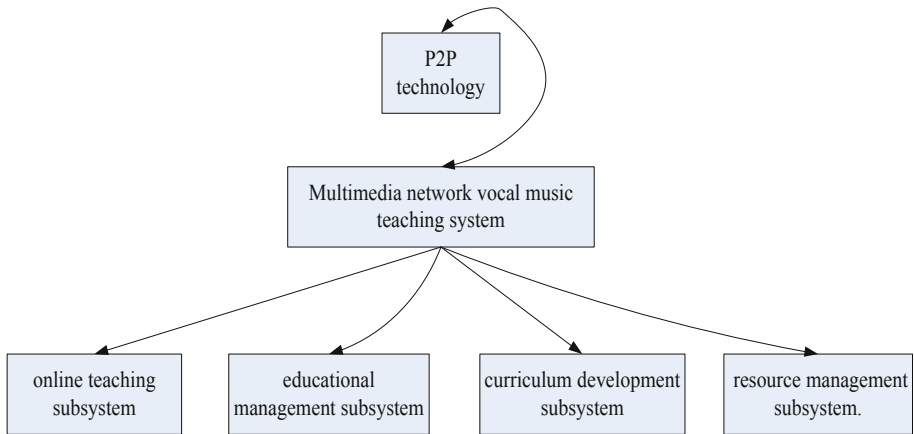


Fig. 1. Multimedia network vocal music teaching system framework

3 Multimedia Vocal Music Learning System Design

3.1 Function Design

Multimedia function of the vocal music study system in the processing of audio system, students can review collection online graphics singing and parameters, the height of students set a specific time, the system will automatically remove storage units required audio data directly, because the audio data is available at any time, and according to the different changing of time, to sound data waveform curve drawing. Students often change their intonation because of the volume of their voice, the flow of air they breathe out, etc. In this respect, the teacher can first sing the deviation of the next piece of data to highlight the deficiency of the students, and then the students can sing to guide the correction, and the students can be in the best mental state to reach the appropriately high volume. This feature can change online statements, and it also has audio replication. An important manifestation of the solid basic skill of vocal music is the precise control of breathing. If students can accurately control their breathing while singing, it will greatly improve the students' sound effect and emotional performance. The time and comfort of breathing have a great influence on the appeal of the song. For the measured values, the function of the standard variance is shown in Formula 1:

$$std(x) = \left(\frac{1}{y-1} \sum_{i=1}^y (x_i - x)^2 \right)^{\frac{1}{2}} \tag{1}$$

3.2 Vocal Music Learning System Construction Needs

On campus, wired Internet is used as communication network, and intelligent multimedia perception information is mainly transmitted through wired Internet. At the same time, the wireless network can also be connected. Since wireless access networks are difficult to distinguish between human and machine access networks, there is no need to create a special wireless access network for the Internet of Things, which can be used as an auxiliary communication network. It can significantly speed up distance learning. Students can get the information they need, such as teaching materials, reference materials and so on, in a very short time through the Internet. There is no time and space limit, students' learning status can be recorded on the Internet, teachers can sum up students' learning status, in order to further improve teaching methods and teaching content. Students can organize classes at their own pace, independent of their own time, place and level, according to their own situation.

3.3 The Convenience of Multimedia Learning Systems

The learning system can combine all the teaching materials, which actually place many components in time and space rather than as an actual combination of documents. Elements can be placed on the same server or on different servers. Different components can be placed on different servers, which not only facilitates maintenance and management, but also the corresponding server space improves the course library according to standards, providing convenient conditions for future parts to place a large number of resources. Promote resource sharing. A course program can be associated with a stream of multimedia files on the network so that more credits can share components' resources. For example, when creating a new teaching materials file, if another teaching unit has already created a video file for the teaching materials, it can add the other unit's video file to its own video elements database with the permission of the other unit without having to do so again. When you choose the type of course, just select this video and you can choose the type of course you want. Because course software can be associated with streaming multimedia documents on the network, streaming multimedia files can be associated with many types of courses. This makes full use of existing resources and avoids duplication of inefficiencies. Textbooks can also be combined with many smaller types of courses, and many types of courses can be constructed. Teachers can combine their teaching experience, depending on the subject, chart, chart, etc. These lessons are easy for students to understand because they are produced by professional teachers. Finally, the intellectual property rights of professionals, teachers and students are protected. Due to the use of streaming media multimedia technology, the data of teaching materials are not stored locally, which clearly hinders replication and dissemination. It basically guarantees the intellectual property rights of natural production units.

For remote multimedia applications that require high real-time performance, streaming media can be continuous and live, from the video frequency server to the user's computer. Instead of waiting for the entire file to load, the client only needs to monitor after a startup delay of a few seconds or 10 s. At the same time, the copy download method saves the storage space of the client and effectively uses the bandwidth. Streaming media technology can be transmitted and copied simultaneously without leaving any

copies of the customer, thus effectively preventing the spread of data and protecting the intellectual property rights of creators.

4 Analysis of Vocal Music Learning System Based on Network Information Technology

4.1 Current Situation of Vocal Music Learning System

In the process of vocal music teaching, the teacher must play the role of multimedia teaching system effectively and guide students to master rich vocal music materials. With the support of multimedia technology, teachers can collect many teaching materials related to sound and music teaching, and teachers can apply these materials in classroom teaching. For example, the content of a lesson must introduce certain Musical Instruments to the students, and the students have no prior contact with these tools. In this case, if the teacher can also make a detailed explanation and description, the students can not really feel like these phonetic performances are useful and effective means. At present, if teachers can provide students with vivid and rich images of Musical Instruments through multimedia teaching facilities and materials that students use, then students can have a deeper understanding of Musical Instruments. Since then, it has not only greatly enriched the content of classroom teaching, but also the students who actively study vocal music themes can be effectively stimulated, and the quality of vocal music teaching can also be improved. In the teaching process, teachers must pay attention to the process of learning innovation in the case of mobilizing students really polar, in order to let students participate in singing and games of the actual activities, really participate in voice music education and activities. Teachers can also guide students to appreciate the process of music, so that students in the musical type of analysis, emotion and work in a way, so that students can not only train pronunciation, but also in the ability to appreciate music in music can also do a great practice. In addition, teachers can design a realistic singing environment for students, singing training in a relatively formal singing environment, so that students realize the importance of singing skills.

By shown in Fig. 2, we can see, the used data analysis during 2008, it doesn't really account for more than learning vocal music, most education way by learning education in the college entrance examination, but with the different development of The Times, way more and more students, more students are development, through the study of vocal music as an art was born at the same time, The demand of multimedia vocal music learning system is increasing. It can be seen that the demand of multimedia vocal music learning system is positively correlated with the proportion of students.

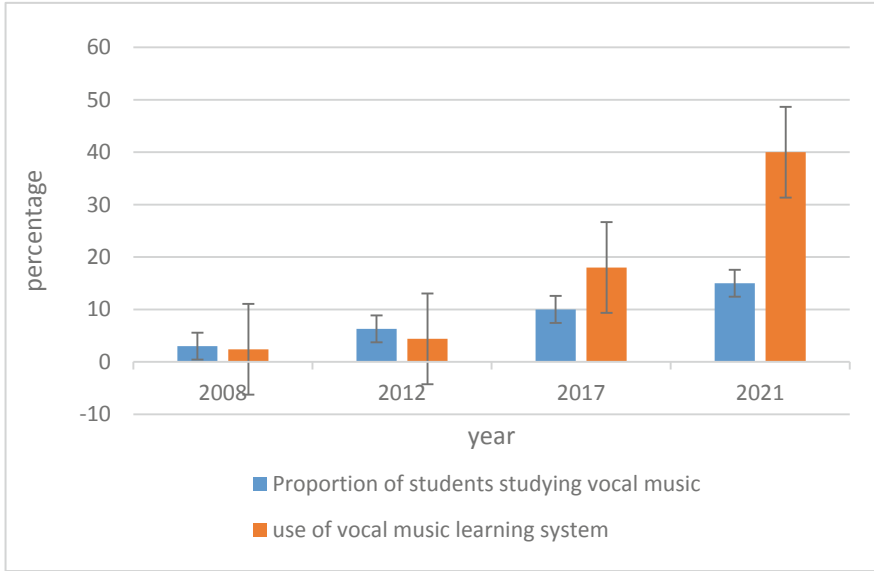


Fig. 2. Proportion of vocal music learning and system from 2008 to 2021

4.2 Streaming Media Improves the Real-Time Performance of Vocal Music Learning System

The network is always unpredictable. While establishing a fixed-size compliance network with customers, the most intractable problem of multimedia network synchronization is offset clocks and so on, they may cause terminals to take saturation or exhaustion, and the effect of multimedia is to jump and jump. So you can't wait for the buffer to actually empty or fill, but be forewarned that once the media stream arrives, it will be necessary to use the callback function to call audio and video for transmission and playback. The network efficiency is calculated as shown in Formula 2:

$$shortest\langle path \rangle(G, source = None, target = None) \tag{2}$$

The data transmission management module sends the required information to the corresponding node according to the information of the node that has arrived. The means of transmission flow are divided into single wave transmission and group transmission. One-way transmission is atomic function, and transmission is single function. With the increasing number of network users, on-demand transmission is a waste of network resources. With the group transfer method, the media server only needs to send one message file, and all clients sending the request share the same message file. Information can be sent to users at any address, thus reducing the amount of information material transferred over the network, which makes using the network more efficient. However, each student may see very different visual frequency data, and even watch the same video, leading to the progress of watching may be different due to different integration time, and may lead to delay, which still needs continuous improvement.

5 Conclusions

In university as long as it can effectively use multimedia teaching system, multimedia network education a new education mode in student use of multimedia technology, computer technology and network technology for learning and information resources to lead the teachers achieve the teaching goal, the goal of learning and effectiveness of the quality of teaching will be adapted to the development of The Times. In the future, teachers should avoid the old ideas and make proper use of the multimedia vocal music teaching system in colleges and universities to train and bring up a group of more musical talents.

References

1. Chen, Y.: Design and implementation of college English learning management system based on mobile learning background. *Revista de la Facultad de Ingenieria* **32**(9), 285–290 (2017)
2. Baldwin, R.: The great convergence: information technology and the new globalization. *Bus. Econ.* **53**(4), 1–3 (2018)
3. Sambamurthy, V., et al.: The design of information technology planning systems for varying organizational contexts. *Eur. J. Inf. Syst.* **2**(1), 23–35 (2017)
4. Hoang, D., Tran, N., Boffo, C., Marinozzi, V., Stoynev, S.E.: Intelliqunch: an adaptive machine learning system for detection of superconducting magnet quenches. *IEEE Trans. Appl. Supercond.* **31**, 1–5 (2021)
5. Amitabh, D.: The role of information technology policies in promoting social and economic development: the case of the State of Andhra Pradesh, India. *Electron. J. Inf. Syst. Dev. Ctries.* **19**(1), 26–33 (2017)
6. Xiong, P., Buffett, S., Iqbal, S., Lamontagne, P., Mamun, M., Molyneaux, H.: Towards a robust and trustworthy machine learning system development (2021)
7. Mocetti, S., Pagnini, M., Sette, E.: Information technology and banking organization. *J. Financ. Serv. Res.* **51**(3), 313–338 (2017)
8. Luftman, J., Lyytinen, K., Zvi, T.B.: Enhancing the measurement of information technology (IT) business alignment and its influence on company performance. *J. Inf. Technol.* **32**(1), 26–46 (2017)
9. Hu, Q., et al.: Control system design and implementation at flexible, distributed offshore sensor test sites in the Yangtze Estuary Area. *J. Ocean Univ. China* **20**(2), 285–295 (2021). <https://doi.org/10.1007/s11802-021-4469-x>
10. Yue, J., Wang, Z.B., Lai, H.B.: Smart gate system design and implementation based on cloud platform. *Procedia Comput. Sci.* **154**(6), 40–46 (2019)



Sports Quality Test Recognition System Based on Fuzzy Clustering Algorithm

Dayang Ming^(✉)

Sports Department, Shenyang Jianzhu University, Shenyang 110168, Liaoning, China

Abstract. Deepening the understanding of the fuzzy clustering algorithm, expanding the scope of application of the fuzzy clustering algorithm, and promoting its integration with sports quality testing will help promote the innovation of the sports quality testing recognition system. The purpose of this article is to study the physical fitness test and recognition system based on fuzzy clustering algorithm. First, the performance evaluation index of fuzzy clustering algorithm and the group visualization method based on Sammon mapping are introduced. The structure of CAPTES system and the function of each subsystem are introduced and analyzed. The system is a highly implemented professional system whose main function is to identify and analyze sports qualities. The fuzzy cluster analysis module makes statistical analysis simple and easy. After the successful development of the sports quality test and recognition system based on the fuzzy clustering algorithm, it was tested in M Academy in time. The experimental results show that the average data of the students with good performance in the shoulder rotation test is 110.6 cm, which is higher than the students with excellent performance in the shoulder rotation test. Therefore, the feasibility of the shoulder rotation test as a test of physical fitness remains to be studied. The scoring test on the data of the first and second grades of M College proved that the software system is stable and effective, and it has been well received by the school's physical education department.

Keywords: Fuzzy clustering · Physical fitness · Test recognition · Recognition system

1 Introduction

The physical condition of adolescents in my country is on a downward trend. Although it is related to many factors, unhealthy lifestyles and insufficient physical exercise are the main reasons [1]. In today's rapid economic development, many diseases caused by unhealthy lifestyles continue to spread. Therefore, cultivating students' healthy lifestyles and exercise habits is the top priority of school work [2]. The majority of scholars also put more energy on how to improve people's sports quality, life quality and life satisfaction through exercise and healthy diet. The main goal of student sports quality research is to promote the improvement and improvement of students' sports quality their lives [3, 4].

Currently, fuzzy clustering algorithms are widely used in recognition systems [5]. The traditional fuzzy C-means clustering algorithm of Li X X has poor anti-noise ability

in image segmentation and poor clustering results. In order to overcome this problem, this paper proposes a new image clustering algorithm based on SLIC superpixels and transfer learning. The algorithm uses the SLIC superpixel method to improve the edge matching degree of image segmentation and enhance the robustness to noise. Transfer learning is used to correct the image segmentation results to further improve the accuracy of image segmentation. In addition, this algorithm improves the original SLIC superpixel algorithm, making the edges of superpixels more accurate. Experimental results show that the algorithm can obtain better segmentation results [6]. Lee SM proposed a moving target extraction and tracking method to improve animal recognition and tracking technology. First, we propose a method of using FCM (Fuzzy C-Means) clustering algorithm to merge separated moving objects into one moving object to solve the problem of moving objects lost in the process of moving object extraction. In addition, we propose a method of extracting data from moving objects and a method of counting moving objects to determine the number of clusters to meet the conditions for performing FCM clustering algorithm [7]. It is feasible to use fuzzy clustering algorithm to establish a physical fitness test and recognition system.

Among the many indicators of student's physical quality, which is the most representative, this issue is also a hot and difficult point of physical quality. This paper uses a sports quality test and identification system based on fuzzy clustering algorithm to verify the weight of each index, and then select the index, and use this certainty to change the quantification, and the method of determining the index is more scientific and logical. On the basis of determining the percentage contribution of the index weight to the sports quality to optimize the index, the mathematical model is established through cluster analysis and discrete analysis to quantitatively evaluate the student's sports quality, which can not only effectively improve the student's quality. The level of movement quality has important practical significance, and also provides some experience and reference for optimizing routes and solving other similar problems.

2 Research on Sports Quality Test Recognition System Based on Fuzzy Clustering Algorithm

2.1 Physical Fitness

Physical fitness is a complex concept composed of "sports" and "quality". The definition of individual physical fitness should be placed in the field of sports, combining the concept of quality and the structural characteristics of quality [8]. This article believes that fitness is the sum of various sports abilities that people have in physical activities, and these abilities are formed through acquired physical learning based on specific innate physiology.

2.2 Fuzzy Clustering Algorithm

There are many performance evaluation indicators for fuzzy clustering algorithms, such as clustering accuracy, clustering efficiency, clustering validity function, noise response ability, etc. [9]. In addition to quantitative evaluation, we can also use the method of

grouping renderings to visually analyze the clustering results of the fuzzy grouping algorithm. Since a set of clustering samples is often a set of high-dimensional data with multiple features, it currently involves data visualization technology [10]. The basic idea of data visualization is to minimize some characteristic high-dimensional data sets through specific transformations, highlight the features that best reflect the topological structure and interaction of the original data set, and map them to low-dimensional space [11]. In this way, multi-dimensional data analysis and research can be carried out in a small space.

2.3 Sports Quality Test Recognition System

(1) System composition of CAPTES

CAPTES is mainly composed of a statistical analysis module and a separate quality evaluation module [12]. Its main function is to perform statistical analysis on exercise data and evaluate individual performance. Among them, statistical analysis mainly uses principal component analysis and cluster analysis to analyze certain problems in teaching and training, and provide a theoretical basis: the personal scoring subsystem is used for personal scoring, that is, sports management. An important part of the main function is to improve the accuracy of the grading process and reduce the workload of teachers.

(2) Data flow in the system

The design of this system takes into account the future development of data acquisition technology, so the data flow method of the data acquisition part is pre-defined in the design part, and language selection is arranged. Mainly choose C and Foxpro as the system design language.

(3) The function of each subsystem

1) System interface

The system interface mainly includes the user interface and the main control of the system. Mainly used to drive and tune the normal operation and tuning of various subsystems: to provide users with a convenient and flexible operating environment.

2) Principal component analysis subsystem

The principal component analysis subsystem consists of the mathematical model of principal component analysis, which mainly analyzes the initial data, finds the principal components (variables), and provides a basis for the analysis of sports quality.

3) Fuzzy cluster analysis subsystem

The fuzzy cluster analysis subsystem is mainly composed of various mathematical models of cluster analysis (the shortest distance method, the intermediate distance method and the deviation square sum method). It is mainly to analyze the physical quality based on the analysis of the main components, and divide the physical quality indicators into different groups according to the categories, so as to provide the basis for teachers to teach according to their ability.

4) Individual evaluation subsystem

The personal assessment subsystem is mainly composed of three parts: standard module, quality and energy module, and classroom scoring module. Its main function is to use it to score students' physical fitness and record the results in the library

to reduce the workload of teachers. The mass-energy unit is the adjustment of the evaluation result of the last unit, and the “score” can be adjusted according to the different physical conditions of the human body. Ketang scores are adjusted specifically for sports performance management. Mainly used to replace labor, calculate and compile performance reports.

3 Investigation and Research on Sports Quality Test Recognition System Based on Fuzzy Clustering Algorithm

3.1 Research Objects

This article randomly selects 12 students from the first and second grades of M College and forms six groups of unit 2 for comparison. Recognize their physical fitness test as the research object. The survey subjects are the on-the-job college physical education teachers in the first and second grades of M College and the sports scientific researcher of M College.

3.2 Test Method

800-m run: At least two people are tested in pairs, starting from standing. Start running after listening to the “start” command. The timer sees the flag and starts timing, and stops when the person’s torso reaches the vertical level of the finish line. Record test results in minutes and seconds, minus decimals.

Standing long jump: Individuals stand with their legs naturally separated. After standing up and jumping, do not press the line on your toes (it is best to use a rope when jumping). Take off with both feet at the same time, without stepping or jumping. Measure the vertical distance from the back end of the jumper to the nearest point. Try three jumps and record the best result. In centimeters, no decimals are calculated.

Throwing a solid ball: In the test, a person stands behind the throwing line, with both feet up and down or left and right, facing the throwing direction, lifts the ball to the top of the head with both hands, then leans back, and then throws the ball forward forcefully on the spot. If the feet are straight forward, the back foot can take a step forward when the ball is thrown, but the line must not be pressed. Each throws 3 times and records the best result between them. The record is in meters, with one decimal place. Measure the vertical distance from the back end of the throwing line to the back end of the ball. In order to accurately measure the results, you must be responsible for observing the position of the solid ball.

Sit-ups: Lie on your back on the mat with your legs slightly open, your knees bent at a 90-degree angle, and your two toes cross and attach to the back of your head. The other partner pressed the subject’s ankle joint to stabilize the subject’s lower limbs. When the subject is sitting, the elbows touch or stretch the knees once. When lying on your back, your shoulders should touch the pillow. When the physical education teacher makes a “start” sound, the counter starts timing and records the number of times completed within 1 min. At the end of 1 min, although the subject sat down, but the elbow joint did not reach the knee, the number of times was not counted, and it was in units.

3.3 Data Preprocessing

The comprehensive development level algorithm of the physical fitness test of the first and second grade students of M College is shown in formula 1:

$$\overline{An} = (A_1 + A_2 + A_3...A_n) \tag{1}$$

In the formula, \overline{An} represents the standard score of the physical quality test of the first and second grade students of M College, and n is one of the 10 tests.

The physical fitness test of the students in the first and second grades of M College is evenly divided, as shown in formula 2:

$$B_n = A_{\max} - A_{\min} \tag{2}$$

In the formula, B_n is the standard score of each quality.

4 Investigation and Research Analysis of Sports Quality Test Recognition System Based on Fuzzy Clustering Algorithm

4.1 Cluster Analysis of Special Sports Quality

In the clustering process, the first column indicates the order of clustering, and the second and third columns refer to the two categories merged in each step. The fourth column refers to the aggregation coefficient. As shown in Table 1, since the relationship coefficient between the 5th and 3rd items is the largest, which is 0.887, the first step is to classify the 5th and 3rd items into one category, and the fourth step is to merge with the 6th category. And so on, and finally merged into one category. The aggregation coefficient according to the grouping order is shown in Fig. 1.

Table 1. Clustering schedule

Stage	Cluster 1	Cluster 2	Coefficients	Next stage
1	5	3	0.887	4
2	2	4	0.829	0
3	3	1	0.801	7
4	5	6	0.762	0
5	8	7	0.744	9
6	1	2	0.720	8
7	3	9	0.712	0
8	1	8	0.705	0
9	8	10	0.701	10
10	8	5	0.681	0

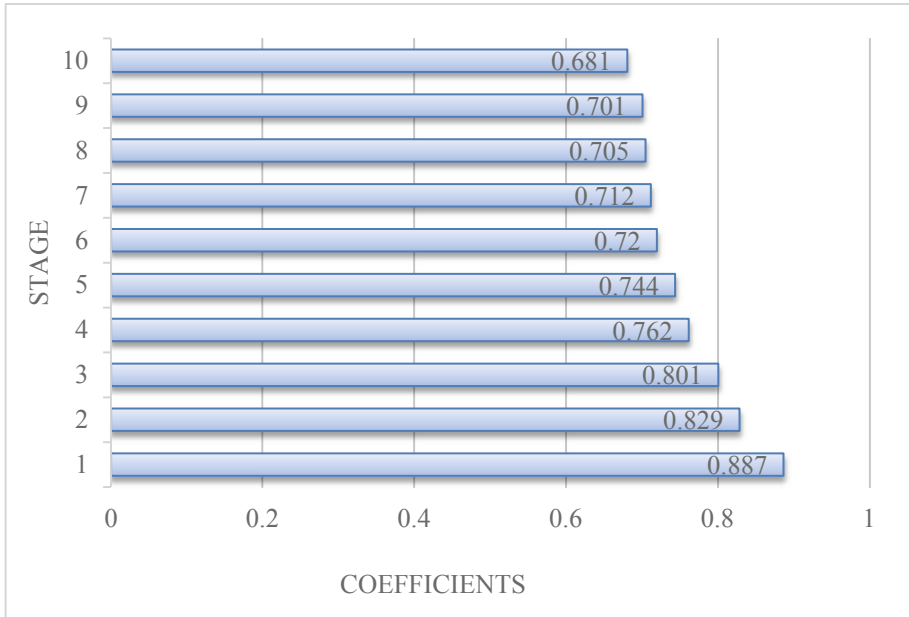


Fig. 1. Coefficient of aggregation

Relatively speaking, in the first category of components, the correlation coefficients between push-ups and sit-ups, standing long jumps, solid ball throws, and skipping ropes are 0.811, 0.754, 0.721, and 0.741 respectively; sit-ups and standing long jumps, solid ball throws, and The correlation coefficients between skipping ropes were 0.742, 0.801, 0.812; the correlation coefficients between standing long jump and solid ball throwing and skipping rope were 0.782 and 0.628 respectively; the correlation coefficients between solid ball throwing and skipping rope were 0.799 respectively. It can be seen that there is a great connection between classes and classes, and it also reminds us that special strength training can greatly improve the physical quality of the first and second students of M Academy. In the second category, the correlation coefficients between the 50-m run and the 100-m run and the retracement run are 0.887 and 0.782 respectively; the correlation coefficient between the 100-m run and the retracement run is 0.780; the third category is only the 800-m run. The correlation coefficient between this sport and other events is mostly around 0.65, but the correlation coefficient of the first group of skipping rope is as high as 0.8, which shows that it has some of the same characteristics as the first group of categories. The fourth type of shoulder has a small correlation coefficient with other indicators, indicating that the fourth type is relatively independent.

4.2 Identification and Analysis of Physical Fitness Test

According to the cluster analysis method, the correlation coefficient (r) was tested, and the results are as follows: The quality of the main sports items of the first and second grade students of M College can be temporarily divided into five categories. The

first category includes only 6 events, namely push-ups, sit-ups, long jump, solid ball throwing, and skipping rope. They all belong to endurance events. Long jump, strong throwing and jumping into the sea are explosive events; and the second category can also be divided into two categories: 50-m running and 100-m running are short-term speed events, and retracement is a test of physical sensitivity. Item; 800-m race belongs to the third category; skipping rope can also be divided into the fourth category; the shoulder rotation test is classified into the fifth category. However, through the data analysis of the shoulder rotation test, it is found that the test data is not closely related to the performance of the students, and most of the data are concentrated between 100–120 cm. Therefore, these six types of items integrate the most detailed identification methods of all sports test indicators, and can reflect the basic characteristics of the physical quality of the first and second grade students of M College (Fig. 2).

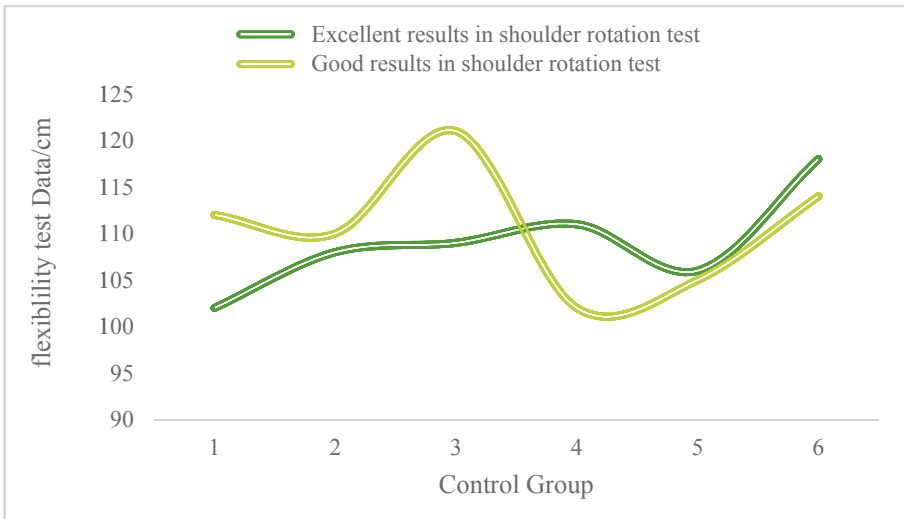


Fig. 2. List of test results of flexibility

5 Conclusions

As the current fitness model is becoming more and more popular with the public, students do not have enough money and time to seek the guidance of professional personal trainers due to many factors such as economy and study pressure. This article is a detailed method of fuzzy grouping. In the grouping process, the algorithm dynamically adjusts the weight of each attribute to each different category according to the specific characteristics of different categories, and fully reflects the difference in the contribution of each attribute to the grouping result. Applying the studied clustering algorithm to the identification system of sports quality testing, it verifies the effectiveness of the fuzzy clustering analysis method in the field of sports quality testing. Improve people's sports enthusiasm and participation in sports activities to achieve the purpose of physical fitness.

References

1. Liu, Z., Jia, H., Wang, Y.: Urban expressway parallel pattern recognition based on intelligent IOT data processing for smart city. *Comput. Commun.* **155**(Apr), 40–47 (2020)
2. Aminanto, M.E., Kim, H.J., Kim, K.M., et al.: Another fuzzy anomaly detection system based on ant clustering algorithm. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **100**(1), 176–183 (2017)
3. Khan, M., Zeeshan, M., Song, S.Z., et al.: Types of complex fuzzy relations with applications in future commission market. *J. Math.* **2021**(4), 1–14 (2021)
4. Jalal, D.K., Ganesan, R., Merline, A.: Fuzzy-C-means clustering based segmentation and CNN-classification for accurate segmentation of lung nodules. *Asian Pac. J. Cancer Prev. APJCP* **18**(7), 1869–1874 (2017)
5. Zhao, Q., Zhang, H., Zhao, X., et al.: Fuzzy clustering image segmentation based on neighborhood constrained Gaussian mixture model. *Pattern Recogn. Artif. Intell.* **30**(3), 214–223 (2017)
6. Li, X.X., Shen, X.J., Chen, H.P., Feng, Y.C.: Image clustering segmentation based on SLIC superpixel and transfer learning. *Pattern Recogn. Image Anal.* **27**(4), 838–845 (2017). <https://doi.org/10.1134/S1054661817040101>
7. Lee, S.M., Seong, I., Joo, Y.H.: Recognition and tracking of moving objects using label-merge method based on fuzzy clustering algorithm. *Trans. Korean Inst. Electr. Eng.* **67**(2), 293–300 (2018)
8. Gao, J., Zhu, Y., Jia, Y.: Pattern recognition of unknown partial discharge based on improved SVDD. *Sci. Meas. Technol. IET* **12**(7), 907–916 (2018)
9. Tang, J., Yu, S., Liu, F., et al.: A hierarchical prediction model for lane-changes based on combination of fuzzy C-means and adaptive neural network. *Expert Syst. Appl.* **130**(SEP), 265–275 (2019)
10. Tan, A., Zhou, G., He, M.: Rapid fine-grained classification of butterflies based on FCM-KM and mask R-CNN fusion. *IEEE Access* **8**, 124722–124733 (2020)
11. Jiang, Z., Li, T., Min, W., et al.: Fuzzy C-means clustering based on weights and gene expression programming. *Pattern Recogn. Lett.* **90**(APR.15), 1–7 (2017)
12. Kim, E.H., Kim, B.Y., Oh, S.K.: Design of digits recognition system based on RBFNNs: a comparative study of pre-processing algorithms. *Trans. Korean Inst. Electr. Eng.* **66**(2), 416–424 (2017)



Conceptual Model of Effectiveness Evaluation of Foreign Language Courses Based on K-means Mean Algorithm

Guilan Chen^(✉)

School of Foreign Languages, Guangdong Technology College,
Zhaoqing 526100, Guangdong, China

Abstract. With the continuous advancement of science and technology, more and more colleges and universities have begun to explore the evaluation of curriculum effectiveness. This has led to the accumulation of a large amount of information data, which often implies some deep-level useful information. The purpose of this article is to study the conceptual model of the effectiveness evaluation of foreign language courses based on the K-means mean algorithm. Firstly, it studies the use of hierarchical clustering method and K-means algorithm to design a conceptual model of foreign language curriculum effectiveness evaluation, which is applied in examples to mine foreign language teaching curriculum index data and analyze various index characteristics. The effective knowledge obtained is used to serve the development of students, optimize education and teaching plans, and improve talent training programs. Based on the constructed evaluation system and the designed foreign language curriculum effectiveness evaluation conceptual model, the scientific, systematic, objective and information evaluation process of curriculum effectiveness evaluation is realized. The experimental results show that foreign language scores are positively correlated with the completeness of the syllabus, and the correlation coefficient is 0.551. The model can dynamically modify specific indicators and the weight of each indicator, in order to meet the needs of universities and colleges to make adjustments and evaluations according to changes in the needs of the times. It can also classify and analyze the effectiveness of courses to achieve certain data mining functions.

Keywords: K-means mean · Foreign language courses · Effectiveness evaluation · Conceptual model

1 Introduction

The evaluation of the effectiveness of foreign language courses is to judge the various indicators of the foreign language courses of the whole school to determine the effectiveness of all aspects of the foreign language courses [1, 2]. Different universities usually formulate standards for different educational purposes, and adjust them to a certain extent according to policy directions and social needs. The evaluation results can be used as a test of the quality of human resource development in colleges and universities,

and can be used for talent selection [3]. With the rapid development of the Internet, various industries have accumulated a large amount of data, and the continuous update of data mining technology has also shown a mature side in some fields. This means that evaluation becomes convenient [4]. However, in the field of education, there are not many examples of the application of K-means method in education. The widely used method of K-means method inevitably becomes a problem in education, and the field of vision in this direction is very broad [5].

In recent years, with the advancement of a large amount of data, many domestic scholars have gradually realized the importance of K-means mean and have actively explored it. Mean regression analysis refers to a method of estimating the underline trend after new data causes a change in the equilibrium state. Jeong-Hwan proposed a new method for interpreting the mirror image of premature ventricular beats (PVC) arrhythmia by applying K-means unsupervised learning algorithm to electrocardiogram (ECG) data. They experimented on the MIT-BIH ECG database. The mirror pattern depicted on the mean recovery HRV data by K-means clustering can be more clearly seen, and the Euclidean metric can be used to identify the difference between normal sinus rhythm and PVC relative distance between cluster centroids And beating [6]. The enhancement method of the k-means algorithm includes calculating a weighted mean to improve the centroid initialization. Fa Bregas AC shows the comparison between the K-Means algorithm and the enhanced KMeans algorithm, proving that the new initial seed selection method is better in terms of mathematical calculation and reliability [7]. Research on the quantification of indicators and the collection and transformation of specific data, as well as specific and feasible indicator system implementation methods, result analysis and feedback, and prediction functions are relatively scarce [8].

The main research innovation of this article is to discover the problems in the course evaluation process and propose solutions through research and analysis. Comprehensive use of the Delphi method and the process of detailed levels, combined with the current situation of university foreign language curriculum construction, participate in the configuration of the university foreign language curriculum evaluation system. Exploring the establishment of a conceptual course evaluation model based on university course evaluation, using structured analysis, combining course statement, grading index system, grading assignments and course design, and creating a high-quality course grading model to study evaluation scores. This conceptual model can be completely replaced The traditional course evaluation method can be used for course evaluation on the road to a paperless office.

2 Research on Conceptual Model of Effectiveness Evaluation of Foreign Language Courses Based on K-means Mean Algorithm

2.1 The Evaluation Index System of the Effectiveness of Foreign Language Courses

Educational conditions are a very important indicator of universities. Most of these indicators are used in practical courses. The first is whether the selection of teaching materials meets the relevant standards, and whether the courses, cases, and reference books are

abundant. The second is whether there are special laboratories, data rooms and examination rooms, whether the equipment is complete, whether each university has a fixed experimental center and the utilization rate of the experimental center [9, 10].

The content of education is directly related to the quality of school management and should be given great attention. Educational content evaluation includes curriculum content, curriculum settings, educational planning and educational applications. It is directly related to the training quality of school staff and is the main content of course evaluation. The effectiveness of education reflects the characteristics of participating in curriculum evaluation at multiple levels and from multiple perspectives [11, 12]. Universities mainly examine the satisfaction of students, professionals, and teachers from the effectiveness of classroom teaching. In some universities, it also includes social awareness and employment growth prospects.

Curriculum features are the characteristics and styles gradually formed in the construction of university courses. Currently, it is highly valued in most universities. Mainly refers to people who have made outstanding contributions to professional courses. Whether it has published high-quality educational research papers or educational research reforms, etc.

The evaluation of teaching methods plays a diagnostic role in the quality of classroom teaching, enabling teachers to quickly find and correct teaching deficiencies in a timely manner. It mainly focuses on the methods and measures used by teachers. In the implementation of classroom education, whether teaching methods and methods meet the needs of students to acquire knowledge, whether new technologies will be introduced, and whether the teaching process emphasizes student-centeredness.

2.2 The Design of a Conceptual Model for the Effectiveness Evaluation of Foreign Language Courses

Its structure and summary are as follows:

- (1) Collect sample data.
- (2) Analyze and process the data in combination with the warning value. The main work of this step is to use statistical knowledge and the index analysis method of foreign language course effectiveness evaluation, combined with the early warning value settings of various places, to judge and eliminate the obvious wrong data in the training sample.
- (3) Construct the training sample set P of the probabilistic neural network on the basis of step two.
- (4) Use random neural network to detect suspicious specimens. This step is to divide the P training sample set created in the third step into two parts. One part is bound as P_1 training samples, and the other part is used as a set of T_1 test samples. By checking the result of T_1 and comparing it with the definition of the T_1 segment of the original training sample P , inconsistent sample points are detected, and a suspicious sample table S is established.
- (5) Use hierarchical grouping technology and K-means algorithm to include a suspicious sample S_1 matrix from N blocks. At the same time, create a central table C of suspicious samples. This step is divided into two steps. First, use the PDIST

command issued by hierarchical grouping to find the distance between each sample point in the training sample set P and each sample point in the suspicious panel S, and find the sample point near each suspicious sample point as a new Suspicious sample matrix S1. Then use the K-means algorithm to edit the table S1, divide the distance table between each sample point according to the level by K, even if S1 produces K clusters. At the same time, create a table C for each cluster center. C contains K vectors.

- (6) Create a new random neural network training sample. In this step, first remove the suspicious samples created in step 4 from P, and then add the matrix sample S1 created in step 5 and the center panel C of each cluster to P for new training, creating a sample PT table.
- (7) Use the newly created random neural network to test the test sample T and compare the test results.
- (8) Update the PT training sample table. In this step, if you want to update the training sample, that is, add new training sample data, use the initial cluster center C table, use the PDIST command to test, and use this limit as a sample with a distance value. This can reduce the influence of incorrect data on model decisions when adding new data.

3 Investigation and Research on Conceptual Model of Effectiveness Evaluation of Foreign Language Courses Based on K-means Mean Algorithm

3.1 Data Collection

The data of this experiment comes from a foreign language school of a university, including data on various indicators of students. According to the determined weight table of each index, the parallel K-Means algorithm is used to carry out the mean cluster analysis. In the experimental data, the first-level indicators of each student are regarded as a dimension, so a five-dimensional vector needs to be determined to represent the situation of each student.

3.2 Data Preprocessing

The mean clustering algorithm mainly divides the data samples by measuring the similarity. The most commonly used similarity between samples is the Euclidean distance, the Mann distance and the Min distance. At present, the most widely used Euclidean distance is shown in Formula 1:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (1)$$

Among them, X_i and X_j represent the data samples in the data set, and X_{ik} and X_{jk} represent the data samples in the k_{th} cluster.

The purpose of mean clustering is to obtain a set of optimal partitions, so that the clusters are as compact as possible and the clusters are as independent as possible, that is, the evaluation function value is the smallest. Mean cluster analysis usually uses the sum of squares of errors as the criterion, as shown in Eq. 2:

$$c_i = \frac{1}{N_i} \sum_{\substack{j=1 \\ x_j \in c_i}}^N x_j \tag{2}$$

Among them, c_i is the mean value of the data objects in the C_i cluster, X_j is the data sample in the C_i class, and N_i represents the number of data objects in the i -th cluster.

4 Investigation and Research Analysis of Conceptual Models for Evaluating the Effectiveness of Foreign Language Courses Based on K-means Mean Algorithm

4.1 The Relationship Between Foreign Language Performance and Curriculum Energy Efficiency Evaluation Indicators

In order to more intuitively analyze the relationship between foreign language scores and curriculum energy efficiency evaluation indicators, this paper selects several typical characteristic attributes as references: the number of foreign language courseware, the completeness of the syllabus, the number of high-quality foreign language teaching research papers published, and the introduction of Number of new technologies, etc. Make a preliminary analysis of the relationship between these five characteristics and performance.

The area chart can visually observe the distribution and general relationship between the foreign language scores and the various characteristics of the curriculum energy efficiency evaluation indicators. Next, by comparing the results of the correlation analysis, you can better understand the impact of the scores on the curriculum energy efficiency, as shown in the Table 1 shown. The number of foreign language courseware in the table is an attribute feature in the data set, which defines the degree of teaching conditions based on the requirements, and the value is a decimal in the interval of [0, 1] (Fig. 1).

Table 1. Correlation analysis

	Number of foreign language courses	Syllabus integrity	Number of research papers	Number of new technologies
Correlation coefficient	0.289	0.551	0.246	0.356

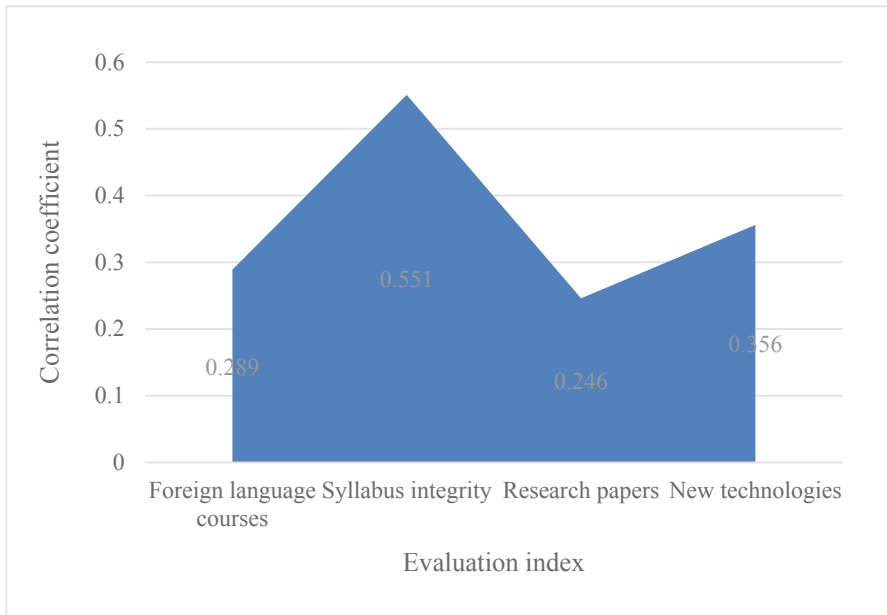


Fig. 1. Correlation analysis

It is not difficult to find that there is a positive correlation between grades and the completeness of the syllabus, and the correlation coefficient is 0.551. The number of foreign language courses, the number of high-quality foreign language teaching research papers that have been published, and the number of new technologies introduced are all slightly positively correlated with the results. If the energy efficiency of foreign language courses is low, it may be difficult for students to improve, and the final academic performance is bound to be greatly reduced.

4.2 Evaluation of the Effectiveness of Foreign Language Courses

About two-thirds of the number of students in the university, the overall oral proficiency is 65 to 80 points, and the range of scores is the largest. Its performance is that all scores can basically meet the standards, but the scores are mostly between 70 and 80. Floating, there is no outstanding performance, a very small number of students are below the passing score as shown in Fig. 2. It can be said that this part of the student situation can basically be used as a summary of the foreign language performance indicators of the school's students, and the basic presentation is the school's overall effectiveness and characteristics of the foreign language curriculum.

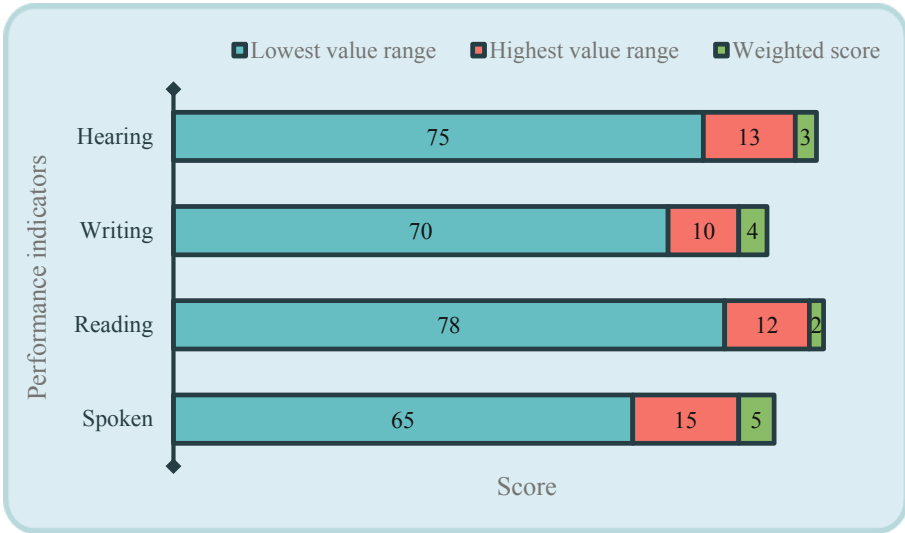


Fig. 2. Foreign language scores are evenly divided and clustered

The student’s reading score is the highest in the index, indicating that the students in this grade pay more attention to the course study, and the grades are better, and they also have a certain performance in reading. However, there are still a few failing students in the three categories of speaking, writing and listening. It indirectly reflects that the syllabus of these parts is not complete enough, or the effectiveness of the curriculum is not obvious enough, the degree of attention needs to be strengthened, and the promotion work needs further development, inspection and adjustment. This kind of students basically exercise and improve their performance in accordance with the school’s syllabus plan, which also indirectly reflects the quality of the school’s foreign language curriculum effectiveness training. They are greatly influenced by counselors, teachers, campus atmosphere and surrounding environment, and they are a group of students who are more obedient and basically follow the educational plan to learn. They clearly recognize and pay attention to the learning of the course. The education management strategy for this kind of students can be guided and strengthened as a whole under the condition of implementing the training plan.

5 Conclusions

Over time, more and more language course index data are stored in the course management center. These data have different educational problems and different educational energy performance in the curriculum. Collect these “typical” questions for analysis and submit student education topics. Carrying out accurate, timely and accurate information analysis is essential for improving and optimizing decision-making, and improving college students’ classroom efficiency. As a non-intrusive technology, mean cluster analysis has a very good market for data mining applications. It can transform complex and hidden data into real decision-making activities and provide a more scientific reference

basis. Therefore, it is very important to study the performance and application of various algorithms. In this paper, a rational analysis of the irregular spatial distribution is carried out, and the overall optimal solution is proposed. By testing experimental data, the K-means algorithm can be used to analyze the energy efficiency of foreign language courses.

Acknowledgments. The Higher Education Project in 2021 of Guangdong Association of Higher Education's 14th five years plan Project's name The reform and exploration of business English major based on 1+x certificate under the big data background Project number: 21GQN69.

References

1. Khaydarova, U., Mustafaeva, N., Abdurakhmonov, B.: Issues on increasing motivation in language learning process. *Int. J. Adv. Sci. Technol.* **29**(5), 1479–1482 (2020)
2. Khong, H.K., Hassan, N.H., Ramli, N.: Motivation and gender differences in learning Spanish as a foreign language in a Malaysian Technical University. *Malays. J. Learn. Instr.* **14**(2), 59–83 (2017)
3. Zhang, Z., Feng, Q., Xu, J., et al.: An approximation algorithm for k-median with priorities. *Science China. Inf. Sci.* **64**(5), 150104 (2021)
4. Golovchun, A., Rakhmanov, N., Abdikalyk, K., et al.: The contribution of world literature's rhymes to Foreign language acquisition. *Man India* **97**(21), 367–379 (2017)
5. Tsalikova, I.K., Pakhotina, S.V.: The issue of motivation for foreign languages learning in international research works: scoping review. *Educ. Sci. J.* **23**(5), 38–63 (2021)
6. J.-H., Kim, Kim, D.-J., et al.: Characterization of premature ventricular contraction by K-means clustering learning algorithm with mean-reverting heart rate variability analysis. *Trans. Korean Inst. Electr. Eng.* **66**(7), 1072–1077 (2017)
7. Fa Bregas, A.C., Gerardo, B.D., Iii, B.: Enhanced initial centroids for K-means algorithm. *Int. J. Inf. Technol. Comput. Sci.* **9**(1), 26–33 (2017)
8. Kanellopoulou, C., Pergantis, M., Konstantinou, N., et al.: Foreign language web-based learning by means of audiovisual interactive activities. *J. Softw. Eng. Appl.* **14**(6), 207–232 (2021)
9. Vasyly, Y., Shyna, N.: The foreign-language communicative competence grounds of master course students. *Sci. Bull. South Ukrainian National Pedagogical Univ named after K D Ushynsky* **2020**(1(130)), 114–119 (2020)
10. Bernstein, J.D., Woosnam, K.M.: Same same but different: distinguishing what it means to teach English as a foreign language within the context of volunteer tourism. *Tourism Manag.* **72**, 427–436 (2019)
11. Cai, L., Huang, H.: Research on distributed data mining technology based on K-mean algorithm. *Revista de la Facultad de Ingenieria* **32**(5), 291–298 (2017)
12. Singh, L., Huang, H., Bordoloi, S., et al.: Exploring simple K-means clustering algorithm for automating segregation of colors in leaf of *Axonopus compressus*: towards maintenance of an urban landscape. *J. Intell. Fuzzy Syst.* **40**(6), 1–25 (2020)



The Acquisition System of Periodontitis Alveolar Bone Structure Based on Digital X-ray Images

Weicheng Yan¹(✉) and Liyun Dai²

¹ Fuzhou Medical College of Nanchang University, Fuzhou 344000, China

² The Sixth People's Hospital of Jiujiang City, Jiujiang 332005, China

Abstract. One of the most important advantages of a digital X-ray intraoral imaging system is its powerful image post-processing capabilities. This is based on such advantages, the digital X-ray intraoral imaging system can be used in many fields of oral science. This article focuses on the periodontitis alveolar bone structure acquisition system based on digital X-ray images. After understanding the relevant theories of digital X-ray imaging and periodontitis alveolar bone structure acquisition on the basis of literature data, it is based on digital X-ray images. The imaging periodontitis trough bone structure acquisition system was tested. Through the stability test of the communication chip of the system, the experimental results showed that the output voltage and current of the chip used in this system have extremely high controllability.

Keywords: Periodontitis · X-ray image · Alveolar bone structure · Acquisition system

1 Introduction

The rapid progress in the application of nuclear technology has become an important force in the innovation and development of new technologies, new materials, and new processes [1, 2]. X-ray spectroscopy is the main tool for quantitative analysis and quantitative analysis of data in materials using X-ray fluorescence [3, 4]. X-ray fluorescence spectroscopy is applied to the determination of mineral trace elements in industrial and mining areas, heavy metals in the atmosphere, drug composition analysis, geological surveys, oil well records, etc. [5, 6]. With the development of electronic manufacturing technology, the energy analysis of the detector has reached a higher level, and the research and optimization of the corresponding electronic back-end measurement system has also achieved certain research results [7, 8].

In response to the application of digitization in the oral cavity, some scientists have used biopsy technology to examine the alveolar bone and identify bone defects before and after treatment, and have achieved positive results. The periodontal ligament is the hard tissue that surrounds the tooth root and connects the alveolar bone to the tooth root. X-ray film shows continuous low linear images with a thickness of 0.15–0.38 mm. If the periodontal ligament is abnormal, X-rays are usually difficult to track. RVG's digital imaging system can produce high-quality images, make local transitions easier

to understand, and reduce dental video image errors [9]. Some scientists use RVG's digital imaging system to measure the height of the alveolar bone. Compared with modern equipment, RVG can provide fast, energy-saving, low-power, high-demand and stable design. When the anatomical model is not clear, the system software can identify the difference of isodensity analysis. The length measurement available at the same time can be used to measure depth and bone density [10]. Some scientists have studied bone density measurement methods through RVG and believe that its application is very effective. Bone density is difficult to measure the quality and richness of X-ray images, but the RVG digital imaging system converts grayscale images into scanned images to improve the resolution of the surgical site. Although macroscopic analysis, grayscale output technology must be used to obtain the standard unit of international speed BMD [11]. To sum up, with the development of digital technology, the application of stomatology has become more and more extensive, but in-depth exploration is still needed in some aspects, so that new digital technology can better serve stomatology.

This article explores the periodontitis alveolar bone structure acquisition system based on digital X-ray images, analyzes periodontitis and the application of digital X-ray images in the diagnosis of periodontitis on the basis of literature data, and then analyzes these theoretical foundations. The periodontitis alveolar bone structure acquisition system based on digital X-ray images is designed, and the designed system is tested, and relevant conclusions can be drawn from the test results.

2 Digital X-ray Imaging and Acquisition of Periodontitis Alveolar Bone Structure

2.1 Overview of Periodontitis

Periodontitis is a chronic disease of the tissues (gum, periodontal ligament, alveolar bone, and cementum) caused by dental plaque. It is one of the most common diseases in stomatology [12]. Basic periodontal treatment is an important step in the treatment of periodontitis, which usually includes oral hygiene training, scanning and subcutaneous scraping. Its purpose is to remove dental plaque, stones, etc., to control or relieve pain, prevent further infections, and plan future treatments. Studies have shown that after treatment, the diseased tissue can be effectively improved. The level of loose teeth may also decrease, but it has not been determined whether treatment can be used to increase alveolar bone.

The alveolar bone supports and stabilizes the teeth, and the teeth are connected to the alveolar bone through the medial ligament. Due to loose teeth, patients with periodontitis often have indigestion. Because the value of alveolar bone has a significant impact on the safety of teeth, monitoring and evaluating changes in alveolar bone can be a reliable basis for diagnosis, treatment and autopsy of patients with chronic diseases. Clinically, the three examinations of epitaxial membrane, inferior membrane and panoramic membrane are only used for the diagnosis and treatment of periodontitis patients. The above three inspection methods are two-dimensional display inspection methods. Since the image is easily enlarged and deformed after imaging, it is difficult to consider the complex anatomy and spatial structure of the oral and maxillofacial region. Accuracy and bone

grade changes are problematic because the above-mentioned clinical methods can clearly simulate the oral cavity or destroy the bone wall.

2.2 Application of Digital X-ray Images in the Diagnosis of Periodontitis

(1) Computer-aided image density analysis technology (CADIA)

CADIA is often combined with DSR technology to measure changes in bone density and bone mass. The CADIA technology is used to quantitatively detect the alveolar bone density of patients undergoing periodontal surgery within 6 months after surgery. In clinical trials, periodontal tissue is used to evaluate the changes in periodontitis after surgery, which can accurately detect the changes in periodontal tissue. It has high application value in periodontal rehabilitation.

(2) Adjustable Aperture Computed Tomography (TACTD)

TACT technology includes optical imaging, digital sensors with CCD sensors for data processing, additional software packages needed to provide additional tomographic functions to the system, and the placement of a diameter opaque ball center area at the X-ray point and target. Every time takes a picture, it has to change the angle to change the imaging, and needs to use multiple X-ray tubes to take pictures on a regular basis. In the final stage, the center of the ball should be determined based on the imaging. The system can recognize three-dimensional tomographic images at any angle and at any depth. The advantage of using this technology for image processing is that it does not require light stabilization and the distance between the object and the light source is large. Because TACT can provide accurate 3D data, clinicians can obtain accurate 3D imaging to identify bone defects before the diagnostic system can function and develop the most appropriate treatment plan for the patient. For example, in patients with three-wall bone marrow transplantation, TACT technology may improve the outcome of bone marrow transplantation. Compared with conventional imaging, direct imaging and TACT are more accurate in detecting diseases after treatment with titanium endotracheal implants. Compared with traditional X-ray film, TACT 3D film can provide more accurate diagnosis and medical information. This helps doctors plan surgeries and cost-effective treatments to improve patient side effects.

(3) Other image processing technologies

Computer image processing systems can also perform image processing by increasing contrast, using edge sharpness and color conversion contour technology. These image processing technologies can play a role in improving the accuracy of digital X-ray diagnosis. After using the silhouette color conversion technology to process digital silhouette images, different levels of gray can be effectively displayed. This is very valuable for clinicians to diagnose and treat.

2.3 Acquisition System Algorithm

According to Sam's law, the optical path design based on the laser triangulation method needs to meet the following two conditions:

- (1) The reflected light from the surface of the object under test can be imaged on the CMOS photosensitive surface;
- (2) When the light bar on the surface of the measured object moves along the incident direction, it can move in a straight line on the CMOS photosensitive surface.

According to the optical imaging formula:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \tag{1}$$

Where f is the focal length of the lens.

$$b = \frac{a \cdot f}{a - f} \tag{2}$$

From the principle of similar triangles, the following formula can be derived:

$$h = \frac{al \sin \alpha \sin \beta}{b + l \cos \beta} \tag{3}$$

In the formula, h : the height relative to the reference plane, l : the image shift distance of the image point, a : the imaged object distance, b : the imaged image distance.

3 Design of Periodontitis Alveolar Bone Structure Acquisition System Based on Digital X-ray Images

3.1 Overall System Architecture

The X-ray spectrum data acquisition system mainly includes two parts: a measuring device and a digital pulse signal processor. The overall block diagram is shown in Fig. 1.

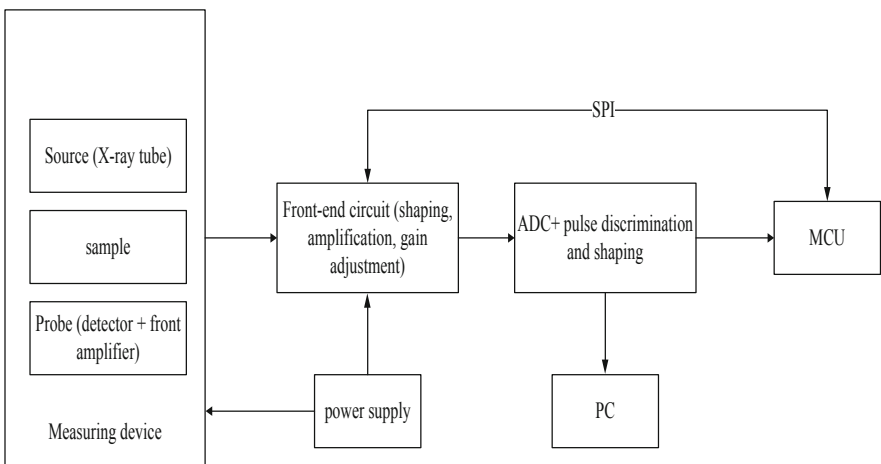


Fig. 1. Frame diagram of X ray spectrum data acquisition system

3.2 The Design of System Hardware

Using the modular design concept, the circuit system is decomposed into design models, which can be divided into control system circuits, receiving circuits, spatial data processing circuits, communication circuits, power circuits, and so on. The control system circuit is the basic function of the receiving system, including a clock circuit, a reset circuit and a debugging interface. The computer is used as the receiving record of the periodontal bone graft function, and the spatial data processing circuit is used to record the spatial data in the image and provide the data to the stitched image. The communication circuit is used to transmit voice data to the host for further data processing. Since the system has different requirements for electrical equipment, the electrical equipment must be modified and processed to meet the requirements of the circuit.

3.3 Acquisition Circuit

First, the laser emits X-rays to irradiate the sample, the sample generates ray fluorescence under the irradiation, and the collector receives the fluorescence. The ray fluorescence interacts with the collector to form electron holes, and moves the current in a differential manner, reaching the collector's two poles for collection before generating a weak current signal. The output of the current signal emitted by the collector is small and weak, so it is not suitable for transmission, so the current signal is transmitted after the preamplifier.

3.4 Spatial Data Acquisition Circuit

In order to synthesize and reconstruct the oral data, it is necessary to record the spatial data of the graph when collecting the oral periodontal data. The system can choose the MPU 3-axis gyroscope to receive the spatial data obtained from the image and convert the measurement signal into a digital output. The various measurement values of the gyroscope can be adjusted, and the user can also select the appropriate detection according to the accuracy of the system. A 1 KB FIFO is integrated in the chip to reduce power consumption. The gyroscope chip can also use the serial port and IIC port for data communication identification, and use the interface identification method to access the following measurement data. The chip adopts digital filtering technology in the international market, which can reduce measurement noise, improve measurement accuracy, and increase the power generation of the system at the same time. This machine is compatible with 3 V/5.4 V power supplies and has simple wiring.

3.5 Communication Circuit

The system uses the communication function to send data from the computer to the host processing system, and uses the PCA96850 chip to convert the FPGA serial port output function to USB communication. The chip is compatible with 5.4 V and 3 V power supplies, the interface supports all dual-port devices, and has temporary storage for sending and receiving data. The maximum data rate is quite reasonable, about 50 bps to 2 Mbps, for large file transfers. This circuit also adds an LED to the TXD interface. When sending data, the input/output status of the LED is determined by the data transmission

bit. This can be used to indicate when to transfer files. Connect the resistance suppressor to the output and receiving ports of the circuit to prevent overvoltage of the digital signal. The purpose of the communication circuit is to convert data output into a communication method to facilitate the connection of computer equipment.

4 System Test

The PWM output precision of the PCA9685 chip is 12 bits, which is converted to a value between 0 and 4096. In the experiment, the initial value of the operating cycle is set to 500, and the output voltage is measured with a multimeter. Then repeat the output voltage measurement until the duty cycle value increases by 500 and the duty cycle value increases to 4000, a total of 8 sets of experimental data. Table 1 shows the output voltage and current values of the multimeter when the working value is 500 and 4000 respectively.

The chip PCA9685 has a 12-bit PWM output accuracy, which is converted to a value of 0–4096. The experiment uses to set the initial value of the duty cycle to 400, and use a multimeter to measure its output voltage. After that, the duty cycle value increases by 400 and repeats the measurement of its output voltage until the duty cycle value is increased to 4000, a total of 8 sets of experimental data. For example, Table 1 shows the corresponding output voltage and current values when the duty ratio values of the multimeter are 400 and 800.

Table 1. System test results

	Voltage (mV)	Current value (mA)
500	500	2.5
1000	1300	5
1500	1900	7.5
2000	2500	10
2500	3100	12.5
3000	3700	15
3500	4400	17.5
4000	4900	20

It can be seen from Fig. 2 that the value of the output PWM duty cycle has a very high linear correlation with the output voltage and current. The square of R of the straight-line adjustment output voltage reaches 0.9999877, and the output voltage and current are highly controlled by adjusting the PWM duty cycle.

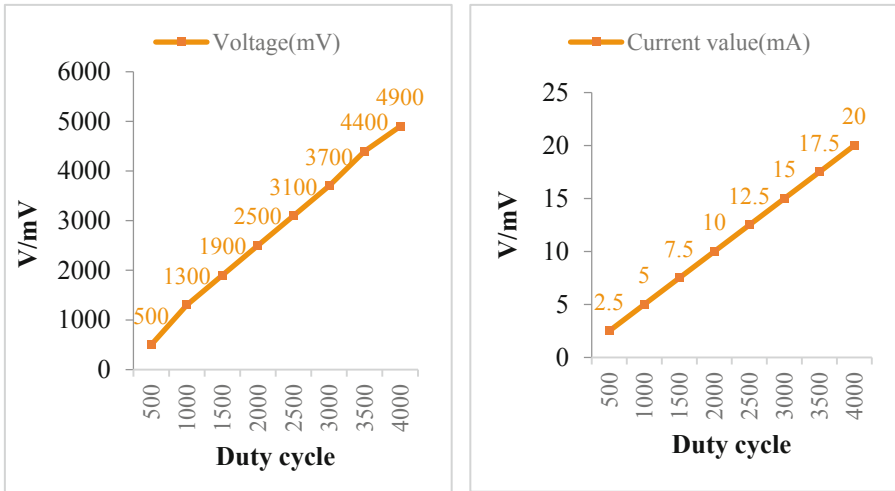


Fig. 2. System test results

5 Conclusions

This paper studies the periodontitis alveolar bone structure acquisition system based on digital X-ray images. After understanding the relevant theories, the periodontitis alveolar bone structure acquisition system based on digital X-ray images is designed and the designed system after testing, it is concluded from the test results that the value of the PWM duty cycle output by the communication chip in this article has a very high linear correlation with the output voltage and current.

References

1. Han, Z.: Clinical analysis of alveolar bone changes after chronic periodontitis based on cone beam CT (CBCT). *J. Med. Imaging Health Inf.* **11**(3), 827–835 (2021)
2. Zhang, Y., Zhang, Y., Zhao, X., et al.: Design and data analysis of sports information acquisition system based on internet of medical things. *IEEE Access*, 1 (2020)
3. Irnstorfer, N., Unger, E., Hojreh, A., et al.: An anthropomorphic phantom representing a prematurely born neonate for digital x-ray imaging using 3D printing: proof of concept and comparison of image quality from different systems. *Sci. Rep.* **9**(1), 14357–14357 (2019)
4. Bhattacharya, O., et al.: Classification of long-bone fractures based on digital-geometric analysis of X-ray images~1. *Pattern Recogn. Image Anal. Adv. Math. Theor. Appl. USSR* **26**(4), 742–757 (2016)
5. Gifford, H.C., Liang, Z., Das, M.: Visual-search observers for assessing tomographic x-ray image quality. *Med. Phys.* **43**(3), 1563–1575 (2016)
6. Favata, J., Shojaee, S.A., Shahbazmohamadi, S.: 3D finite element simulation from non-destructive X-ray tomography and verification with novel mechanical testing and digital image correlation in-situ of focused beam microscope. *Microsc. Microanal.* **26**(S2), 3104–3108 (2020)

7. Skarzynski, L., Suchorzewski, J.: Mechanical and fracture properties of concrete reinforced with recycled and industrial steel fibers using Digital Image Correlation technique and X-ray micro computed tomography. *Constr. Build. Mater.* **183**(20), 283–299 (2018)
8. Lee, K.I., Jang, J.S., Lee, T.R.: Using the X-ray image, augmented reality based electrocardiogram measurement system development. *J. Digit. Convergence* **14**(9), 331–339 (2016)
9. Domenico, G.D., Cardarelli, P., Contillo, A., et al.: X-ray focal spot reconstruction by circular penumbra analysis—application to digital radiography systems. *Med. Phys.* **43**(1), 294–302 (2016)
10. Choi, J.K., Kim, K.B., Kim, G.J.: Image selection algorithm proposal for digital radiography training simulator. *Wireless Pers. Commun.* **89**(3), 833–845 (2016). <https://doi.org/10.1007/s11277-016-3247-3>
11. Bandyopadhyay, O., Chanda, B., Bhattacharya, B.B.: Automatic segmentation of bones in X-ray images based on entropy measure. *Int. J. Image Graph.* **16**(01), 1650001–1650032 (2016)
12. Lee, K.L., Ireland, T.A., Bernardo, M.: Benchmarking the performance of fixed-image receptor digital radiographic systems part 1: a novel method for image quality analysis. *Australas. Phys. Eng. Sci. Med.* **39**(2), 453–462 (2016). <https://doi.org/10.1007/s13246-016-0440-3>



Substation Construction Management and Control System Based on 3D Modeling Technology

Hongwei Yang, Wuchen Zhang, Yamin Wang^(✉), Jizhi Su, and Nan Cheng

Economic and Technological Research Institute, State Grid Hebei Electric Power Co., Ltd.,
Shijiazhuang, China

Abstract. Research on clean building management based on three-dimensional design technologies needs to be considered in the new era of construction and management of transmission and power projects. As the size of the network increases and the size of each project increases, it becomes more difficult to manage the project website and the project information becomes more complicated. High requirements are identified at the level of power transmission system management and transmission construction projects. China's construction management has not complied with the regulations of a large-scale network construction project. Therefore, this document discusses building management and control systems based on 3D modeling technology by analyzing system requirements in the relevant documents. And then develop the building structure management and management system according to the needs of the system to carry out the construction of control stations and management system, then design management. The test results show that, although the system still has its shortcomings, the total number of people with excellent ranking of system performance is more than 32% and more than 30%.

Keywords: 3D modeling · Substation construction · Management and control system · Bim 4D technology

1 Introduction

As an important process of power grid construction, the management and control of substation construction are the key factors that determine the quality of substations [1, 2]. Therefore, our country attaches great importance to the supervision of construction quality [3, 4]. According to the current supervision system, the substation must be strictly supervised from the beginning of construction to delivery to ensure the quality of construction [5, 6]. Although the project must be reviewed in accordance with the regulations before delivery, the quality has not been significantly improved. The important reason is that the supervision and enforcement is not in place and the implementation of the rules is not in place. All researchers have studied these topics [7, 8].

Regarding the research on the substation construction management and control system based on 3D modeling technology, some researchers have proposed that in recent

years, with the continuous development of substations, the task of substation engineering construction has increased sharply, the number of engineering projects has increased, and the corresponding data management work has also been increasing [9, 10]. At present, the management of these documents by the substation staff is still at the stage of traditional manual file management, that is, the corresponding information is first bound into a book in the form of paper, then sorted and organized, and finally submitted to the archives [11]. The current society is making great strides towards informatization. From the perspective of informatization, system informatization should be an application of regional construction and management. It promotes regional management and the development and speed of high-tech technologies through the digitization of information [12].

This paper carries out research on the application of refined construction management of power transmission and transformation projects based on three-dimensional design technology, applying the coordination, visualization, simulation, and optimization functions of three-dimensional design technology, using it in three-dimensional simulation, visual inspection, accurate calculation of engineering quantities, and more accurate, timely and appropriate project management information exchange, etc., make it easier to achieve project management by objectives.

2 Research on Substation Construction Management and Control

2.1 Requirements for Substation Construction Management and Control System

- (1) During the entire project implementation process, use the BIM 4D construction schedule management platform to realize real-time data monitoring and information collaborative sharing, reduce the number of decision-making information transmissions, realize accurate planning, tracking and control, optimize the allocation and organization of various construction resources, and track in real time the progress of the project, timely analysis of the degree of impact of deviations on the overall construction period and important control points and the causes, to achieve real-time control and timely optimization of the project schedule, and to ensure that the project is completed on time.
- (2) Based on the three-dimensional model data, according to the project information management methodology, sort out the various professions/types of data in the construction management process; unified data acquisition, update, storage and transmission standards;
- (3) Based on BIM 4D standardization of grid engineering construction schedule planning and dynamic control methods, establish the entire construction process control with schedule planning as the main line. According to construction project management theory, according to the characteristics of power grid engineering, research the whole process of construction control work flow, establish construction schedule management model, study schedule management early warning mechanism and model, and propose methods to control and adjust deviations in different stages; realize according to construction state, quickly and accurately put forward corrective measures and adjustment plans, and provide data support and decision-making basis for project management personnel for dynamic management of the construction process.

3 Design of Substation Construction Management and Control System Based on 3D Modeling Technology

3.1 The Overall Functional Architecture of the System

Based on the user requirements, key functional requirements and technical requirements described above, and absorbing my country's current practical experience in construction quality supervision, this research establishes a functional model of the construction quality supervision system to support on-site construction quality acceptance management. This functional model is supported by the BIM data management platform and includes 5 modules including system management module, project management module, plan management module, process management module, and result management module, as shown in Fig. 1:

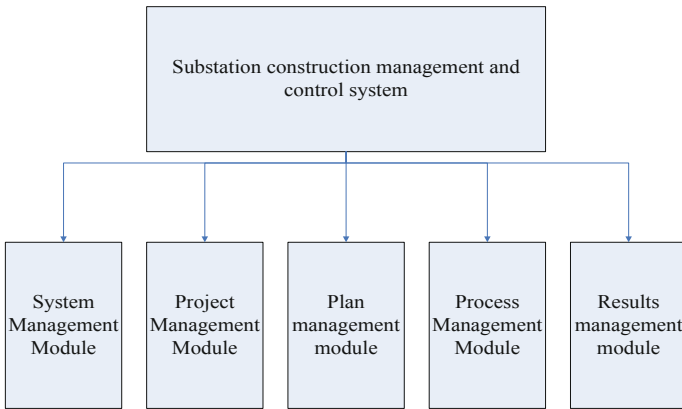


Fig. 1. System total function architecture diagram

3.2 Information Model of Construction Quality Supervision Process

The construction party and the supervising party check the check points under each inspection item at the construction site, and conduct self-inspection and re-inspection respectively for the check points, fill in the corresponding customized construction quality acceptance original record form, and the system automatically judges the check point according to the content of the form, the inspection items and inspection batches are qualified. If they are not qualified, the construction party will be prompted to make corrections and fill in the correction records. The filled-in original record form will be saved to the database for automatic generation of various inspection lot quality acceptance forms. The item information, inspection lot information, inspection item information, inspection point information, inspection records, and inspection results in the form can all be sourced from the site the data is extracted from the record table, so that no secondary transcription is required.

3.3 Information Model of Construction Quality Supervision Results

Associating construction quality result information, including acceptance records, rectification records, and media data, with IFC data will help visually view the construction quality result information and support the use of data in other construction links such as the building envelope stage.

3.4 Inspection Lot Quality Acceptance Workflow

Inspection batch acceptance is the basis and key of on-site construction quality and construction supervision. Due to the complex specifications and numerous procedures involved, it is very important to establish a reasonable work process. For inspectors, they can focus on light data entry tasks such as judgment, inputting values, and sending reports on the spot according to the intelligent prompts of the system, which not only reduces the workload of inspectors, but also organizes data quickly and effectively.

- (1) The first stage: self-inspection by the constructor. At this stage, the inspector of the construction party checks and selects the points to be inspected, and submits the original record form after quality acceptance. The system will automatically determine the inspection result. When the inspection is qualified, the system will notify the supervisor to conduct a re-inspection, otherwise the construction party will rework until the inspection is qualified.
- (2) The second stage: re-inspection by the supervisor. After checking and selecting the points to be inspected, the inspectors of the supervision party fill in the original quality acceptance record form and submit it. The system will automatically determine the inspection result. When the inspection is qualified, the process inspection ends. When the inspection is unqualified, the inspector of the supervision party will send a rectification notice to the construction party.

3.5 Integrated Method of Mobile Positioning Technology

First, establish a WIFI fingerprint database and a geomagnetic signal fingerprint database respectively, and store the data through the data transmission and storage system. In actual positioning, the current geomagnetic data and WIFI intensity information are collected according to the movement of the person, the RSSI positioning algorithm is matched with the WIFI fingerprint in the data storage system, the floor information and rough location information are first obtained, and then the geomagnetic positioning algorithm is used to match the database. In the geomagnetic database, the final position information is obtained. Finally, the positioning information is collected and stored by the network terminal.

4 System Inspection

4.1 Experimental Design

The selected trial object is the 110 kV Yuxiang Substation. Yuxiang Station is the first urban smart benchmark substation in Shijiazhuang urban area and the first fully open

integrated energy substation in the urban area. Yuxiang Station is located on the south side of Nanjiao Passenger Station in Yuhua District, With a total construction area of 2171.8 m². In this phase, two 50 MVA main transformers, two 110 kV incoming lines, 28 10 kV outgoing lines are to be constructed, and the total length of cables to be laid is 2.25 km. The project uses a substation construction management and control system based on Bentley Synchro BIM 4D 3D modeling technology to manage the construction process, and prepares 1075 construction plans and corresponding resource input plans such as man, material, and machinery, as shown in Fig. 2:

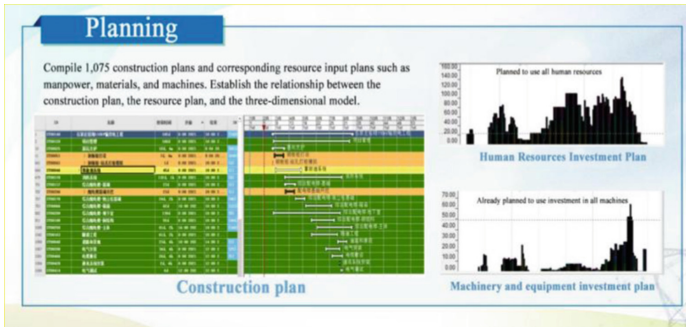


Fig. 2. Resource investment plan based on Bentley Synchro BIM 4D 3D modeling technology

At the same time, the construction simulation of earth excavation, box culvert construction, steel structure construction, and exterior wall installation construction was carried out through the substation construction management and control system of three-dimensional modeling technology, as shown in Fig. 3:

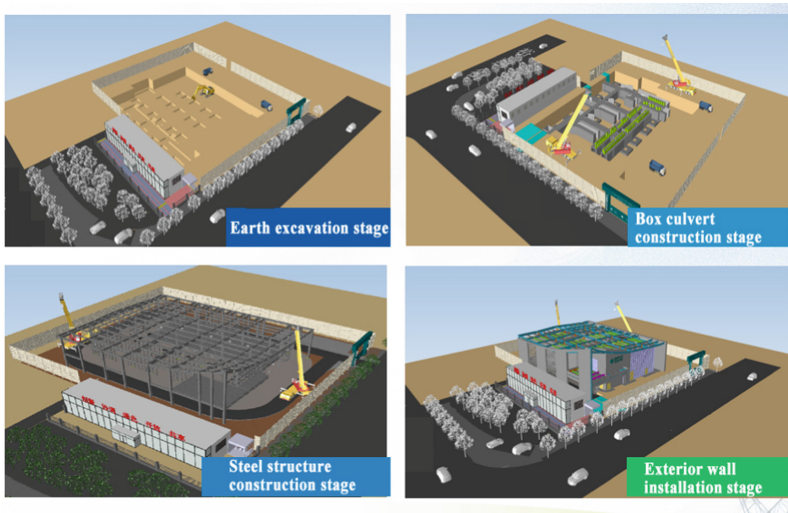


Fig. 3. Construction simulation of 3D modeling technology

4.2 Data Processing

(1) The basic meaning of the Relationship Analysis Method is often to identify a specific relationship between different objective revelations, but they do not fully complement each other on the scale. There are two main ways to identify the objective characteristics of an event: qualitative analysis and quantitative analysis. The main purpose of quality analysis is to rely on the theoretical knowledge of the researcher and the practical experience of the researcher to determine whether the objective relationship between events is different or not. Or whatever the topic, the topic of this method of questioning is pretty strong. Among these is the formula for the usual calculation:

$$r = \frac{S \wedge 2 \ xy}{S_x \ S_y} = \frac{\sum (x-\bar{x})(y-\bar{y})/n}{\sqrt{\sum (x-\bar{x}) \wedge 2/n} \sqrt{\sum (y-\bar{y}) \wedge 2/n}} \tag{1}$$

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x \wedge 2 - (\sum \bar{x}) \wedge 2} \sqrt{n \sum y \wedge 2 - (\sum \bar{y}) \wedge 2}} \tag{2}$$

4.3 Analysis of Experimental Results

The questionnaire examines the control and control system of the substation structure, based on the 3D modeling technology developed in this work, and evaluates the logic of the system. The test results are shown in Table 1:

Table 1. The rationality of construction management and control system of substation based on 3d modeling technology

	System interface effect	Easy to operate	System response time	The content design of
Good	46%	34%	48%	32%
General	32%	45%	30%	47%
Not good	10%	9%	13%	12%
Do not know	12%	12%	9%	9%

Figure 4 shows that although the procedure is still inaccurate, in general, the number of test subjects shows more than 32% and overall it is more than 30%.

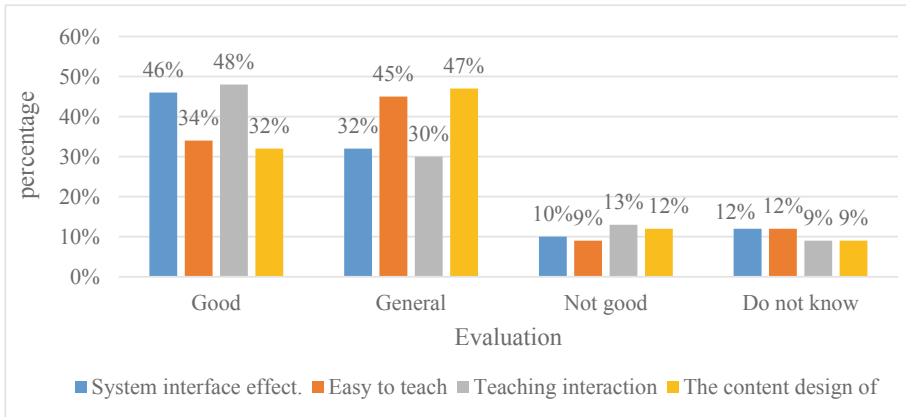


Fig. 4. The rationality of construction management and control system of substation based on 3d modeling technology

5 Conclusions

This document discusses substation construction management and control systems based on 3D modeling technology, provides advice on issues related to construction management and project control, analyzes system requirements for problems and provides practical ideas for the future. Hopefully this article can help to further improve the management and control of substation construction through the management and control system of station construction. In the preparation of this article, it was found that the management and control of the substation construction has not yet been fully formed. It's easy to miss each other because of the many sources of information available on the Internet. However, due to limited knowledge, it may not be deep enough or partially. We will continue to explore current issues in our work and research.

References

1. Ma, A., Chen, L., Li, Z., et al.: The configuration file management and control system construction of smart substation. *Smart Grid* **07**(5), 307–312 (2017)
2. Ahmadian, S., Ebrahimi, S., Malki, H.: Cyber-security enhancement of smart grid's substation using object's distance estimation in surveillance cameras. In: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC). IEEE (2021)
3. Peng, H., Gao, C., Xiong, X., et al.: Research on intelligent periodic test operating system for bay level equipment of smart substation. *Dianli Xitong Baohu yu Kongzhi/Power Syst. Prot. Control* **45**(2), 99–104 (2017)
4. Li, G., Huang, J., Ni, C., et al.: Discussion on the construction scheme of intelligent substation on-site protection simulation test system. *Dianli Xitong Baohu yu Kongzhi/Power Syst. Prot. Control* **46**(17), 152–157 (2018)
5. Zhu, W., Kang, Z., Li, Z., Chen, C., Peng, W.: Research on the construction of communication network topology model of smart substation. In: 2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)

6. Lips, M., Anderson, E., Nishida, K., Schneider, G., Zic, J., Sanders, C., et al.: Reflection on the proposed changes to dose quantities—an industrial perspective. *J. Radiol. Prot.* **41**(4), 1410–1419 (2021)
7. Faschang, M., Cejka, S., Stefan, M., et al.: Provisioning, deployment, and operation of smart grid applications on substation level: bringing future smart grid functionality to power distribution grids. *Comput. Sci.* **32**(1–2), 117–130 (2017)
8. Zhu, M.X., Wang, Y.B., Liu, Q., et al.: Localization of multiple partial discharge sources in air-insulated substation using probability-based algorithm. *IEEE Trans. Dielectr. Electr. Insul.* **24**(1), 157–166 (2017)
9. Moreira, N., Molina, E., Lazaro, J., et al.: Cyber-security in substation automation systems. *Renew. Sustain. Energy Rev.* **54**, 1552–1562 (2016)
10. Gampa, S.R., Das, D.: Optimum placement of shunt capacitors in a radial distribution system for substation power factor improvement using fuzzy GA method. *Int. J. Electr. Power Energy Syst.* **77**(114), 314–326 (2016)
11. Memik, E., Nikolic, S.: The virtual reality electrical substation field trip: exploring student perceptions and cognitive learning (2021)
12. Tulabing, R., Yin, R., Deforest, N., et al.: Modeling study on flexible load's demand response potentials for providing ancillary services at the substation level. *Electr. Power Syst. Res.* **140**, 240–252 (2016)



Design and Implementation of Oral English Test System Based on Data Mining Algorithm

Xiaolan Guo¹(✉) and Neha Jain²

¹ Foreign Language Department, Ganzhou Teachers College, Ganzhou 341000, China

² Chitkara University, Rajpura, Punjab, India

Abstract. With the diversification of spoken English tests and the steady increase in the number of people taking the test, the organization before the test and the grading work after the test have become more complicated. The various low-efficiency and high-cost burdens of the traditional examination model are becoming increasingly apparent. The spoken English test system designed in this paper based on data mining algorithm can solve the problems of high cost and time-consuming of traditional test mode. Among the 160 students, 15% have oral test scores of 40–50, 55.625% have scores of 30–40, 23.125% have scores of 15–30, and students with scores below 15 accounted for 6.25%, indicating that the performance of the system is good when reflecting the test results.

Keywords: Data mining algorithm · Oral English test · System design · Traditional test mode

1 Introduction

With the rapid development of modern computer information technology, computer technology has gradually deepened in various fields, which has also produced fundamental changes in the field of education. Modern teaching and learning methods are constantly innovating, and examinations are an important part of evaluating teaching results, and their information construction and development are relatively lagging behind. Most oral examinations still pass manual supervision and evaluation methods. Therefore, it is very necessary to design an oral English test system for oral proficiency test, which can realize the intelligence of the whole process of the test.

Many scholars have studied the design and implementation of the oral English examination system based on data mining algorithms, but they have obtained good research results. For example, a higher education institution formed an intelligent online examination model with examination paper management as the core. This test mode has created the mileage of computer information test. A series of functions such as test paper question management system, metadata test management system, intelligent review management system are integrated in the system, and advanced computer management methods are realized. The forefront of the technology examination platform [1]. A scholar's research on the application of data mining theory and technology in the field of education mainly focuses on the prediction of test scores, the exploration of the relationship between scores

and courses, teaching evaluation, paperless test system question banks, algorithm applications, etc., but for exam registration data mining research on analysis, score analysis and test paper quality analysis is less [2]. Although the research progress in the design and implementation of the oral English examination system based on data mining algorithms is quite good, it is still necessary to optimize the problems of the system to realize the intelligent examination and reduce the workload of organizing the examination.

This article introduces several commonly used data mining methods. According to the feasibility of the system, a system for the oral English test is designed. The performance of the test system verifies the practicability of the system in the oral test process. In the oral test system this system can be introduced to strengthen exam management and reduce the burden of the teacher from preparing questions to scoring the entire exam process.

2 Discussion

2.1 Common Methods of Data Mining

(1) Decision tree method

The decision tree method is a tree structure, which is very similar to a flowchart. It is studied for classification and prediction. The main purpose is to assign it to many pre-set categories, and at the same time use the addition of a field to note the category of the record information. The root node is preferred. The root node uses a preset practice set, also called a decision node. The nodes in the tree are regarded as attributes of a new record, the branches are regarded as test results, and the categories are reflected by leaf nodes. The record information contained in the exercise set has simple fields for the remark category. Search for each node field, and use the information gain to find the field with the largest amount of information, assign and remark, and then refer to the information in these fields to build a branch system and corresponding sub-systems. The nodes of the set are constructed layer by layer. When the leaf nodes are reached, that is, the types of all records in the node are the same. At this time, the algorithm is terminated and a relatively complete decision tree is constructed [3, 4].

(2) Neural network method

Connect large and simple neurons in a node manner according to a certain rule, and the resulting network system is called a neural network. It is close to the structure and ability of the human brain. It responds dynamically to externally input information to process the information, selects an algorithm to learn from the exercise samples, and then stores the information obtained from the exercise, and puts it in the right of connection in an orderly manner [5].

(3) Rough set method

Rough set is a soft computer technology. According to theory, uncertain information can be processed. Even if a certain characteristic of the information is not quantitatively described, it can also be combined with the algorithm to form the basic law of the problem, and it has a strong ability to deal with this unstable and wrong information [6].

2.2 Demand Analysis of the Oral English Test System

(1) Economic feasibility analysis

The target users of the oral English test system are test administrators, oral teachers and candidates. If a school or institution wants to improve the efficiency of test management, it must improve the efficiency of related test management business processing, simplify the business processing process in the test business process, and enable the test staff, oral teachers and test takers to obtain greater benefits in the entire test business convenient. This system does not have high requirements on the economy, as long as it has a computer and suitable development software, it can be developed. Once the system is developed, it will speed up data processing and save manpower and material resources. Moreover, the standardization of the system's long-term needs also improves the usability and security of the data, and can query all the information of the examination process more quickly, and refine the management of the examination work [7, 8]. In the long run, the organization cost, time cost, and personnel cost of examinations in schools and educational institutions are much higher than the cost of system development. Therefore, the spoken language system is economically feasible.

(2) Operational feasibility analysis

To adapt to the processing work of the oral examination business, the system requires a simple and clear operation process so that the operator can input the examination information accurately, and the input data information is processed through the background server. The system is characterized by very simple operation. The operator can easily organize, manage and confirm the examination work after a short training, and it also facilitates the maintenance of the system. Therefore, advanced technology is used in the design to provide more practical technical support for the ease of use of the oral examination system [9].

(3) Management feasibility analysis

With the rapid development of informatization today, most industries have independent information management systems, but many units have not reached the desired level of informatization management requirements in terms of informatization. At present, examination management is basically done by hand in handling business, and there are many problems in the process of various questions, arrangement, and marking. Such as frequent scoring errors, long business processing time, etc. At the same time, it also increases the labor intensity of test managers and scoring teachers, which brings inconvenience to decision-making analysis for many senior managers of schools and educational examination institutions, and cannot quickly adapt to business changes [10]. Therefore, the construction of the network oral examination system has a more feasible meaning in management.

(4) Policy feasibility

This system can be used for college oral English exams. For example, the oral exams of Band 4 and 6 are nationally regulated exams, and there are no restrictions on laws and policies.

(5) Technical feasibility

The technical feasibility analysis mainly analyzes whether the current technology can realize the function to meet the user's needs, and whether it can be technically realized can be considered from the two aspects of software and hardware. In terms of software, the Microsoft database management system is adopted and the C language is used for development. In terms of hardware currently used in many examination rooms, professional servers and client computers are getting higher and higher technical performance. CPU, memory, hard disks are constantly being upgraded, increasing safety and reliability, and reducing price costs. From the school's point of view, each school has its own professional server and computer room, hardware is no longer a restriction issue [11, 12].

2.3 Fisher's Discriminant Method

The discriminant method is to classify the existing training set according to the discriminant function, so as to determine the relationship between the sub-type target variable and the numerical predictor variable. The basic idea of the Fisher discriminant method is to first project the original m -dimensional X space sample into the p ($p \leq m$)-dimensional Y space, then the discriminant function is the linear function form of the input variable, which is:

$$y = \mu_1 x_1 + \mu_2 x_2 + \dots + \mu_m x_m \quad (1)$$

Among them, $\mu_1, \mu_2, \dots, \mu_m$ represents the discriminant coefficient, which refers to the influence of each predictor variable on y , and the dependent variable y refers to a certain dimension in the sample Y space. Usually, there are many discriminant functions, and m y values can be obtained. In order to obtain the ideal classification and discrimination effect, it is necessary to distinguish samples from different categories as much as possible, and to distinguish the projection direction as much as possible. Therefore, the smaller the sum of squared deviations between samples of each category, the better. When discriminating, the basic step is to first find the center point in Y space, then for a new sample X , use the above formula to calculate its discriminant function, and then calculate the distance between the sample X and the center points of the Y space, and then use The distance discrimination method judges the category to which the sample X belongs. which is:

$$W(Y) = (Y - \bar{Y}) \sum^{-1} (\bar{Y}^{(i)} - \bar{Y}^{(j)}) \quad (2)$$

$$\bar{Y} = \frac{1}{2}(\bar{Y}^{(i)} + \bar{Y}^{(j)}) \quad (3)$$

When $W(Y) > 0$, the new sample X belongs to the i -th category.

3 Oral English Test System Based on Data Mining Algorithm

3.1 System Function Design

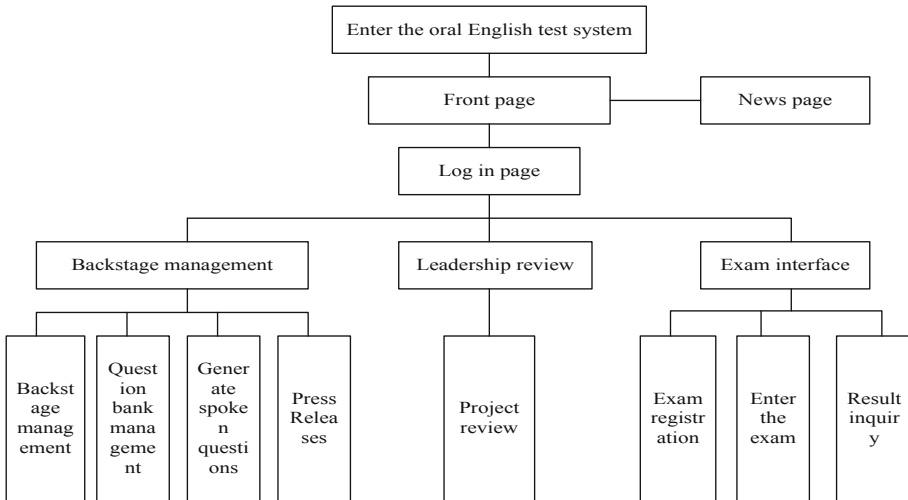


Fig. 1. System function design

Shown in Fig. 1 is the system functional structure diagram. After entering the oral English test system, there will be two pages, one for the home page and one for the news page. Click on the home page to enter the user login page, which includes three modules. The first is the back-end management module. The back-end needs to divide the students of each grade for grade management. The oral exam question bank is also stored in the back-end system. When the exam is approaching, the back-end will automatically generate oral exam question types according to the content learned. There is also a back-end system. The function is to publish exam-related information. The second is the leadership review module, such as the project review of organizing oral examinations. The third is the examination interface, including examination registration, entrance examination and result inquiry.

3.2 User Structure Design of the System

Figure 2 shows the user structure diagram of the oral English test system. There are three types of users in the entire system, namely students, teachers, and system administrators. Students only need to log in to the system to conduct oral examinations. After logging in to the system, teachers can monitor the examination process, review oral question types, and evaluate students’ oral examination scores. The system administrator mainly manages English courses, user logins, and examination information.

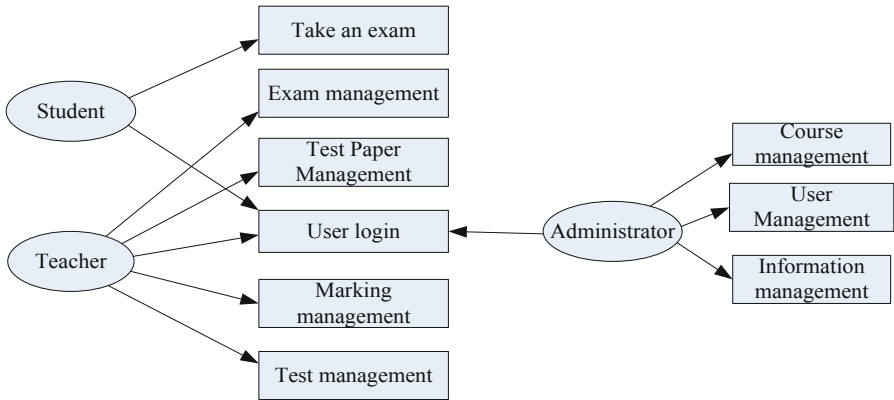


Fig. 2. System user structure design

4 Implementation of an Oral English Test System Based on Data Mining Algorithms

4.1 System Test

Table 1. System test results

Investigation	Target value	Actual value	Pass or not
Response time	<=2 s	1.351	Yes
Concurrency	20	20	Yes
Business success rate	>98%	100%	Yes
Memory usage	<55%	46%	Yes
CPU usage	<50	34.2%	Yes

Connect the designed system to the Internet for testing. The test items have time for system launch, maximum concurrent number, business success rate, memory usage and CPU usage. The test results are shown in Table 1. The reaction time of the system is the time to start the system. The target value should be less than or equal to 2 s. The actual test value only took 1.351 s; the number of concurrent users is the maximum number of users when users use the system at the same time. The target value of the maximum number of users is 20. The maximum number of users of the value is also 20; the business success rate means that the entire process is going smoothly when the system is used for the test. The target value should be greater than 98%, and the test value has reached

100%, indicating that there are no errors in each test process; memory usage refers to the ratio of the system installed on the computer to the computer’s memory, the target value should be less than 55%, and the actual test value is 46%; CPU usage refers to the ratio of the test system to the computer’s CPU, and the target value should be less than 50%, the actual test value is 34.2%. The five test results are all within the test range and meet the design requirements. Therefore, the system can be used normally in the real oral test.

4.2 Analysis of System Application Results

Table 2. Examination results statistics table

	Number of people	Percentage (%)
A	24	15
B	89	55.625
C	37	23.125
D	10	6.25

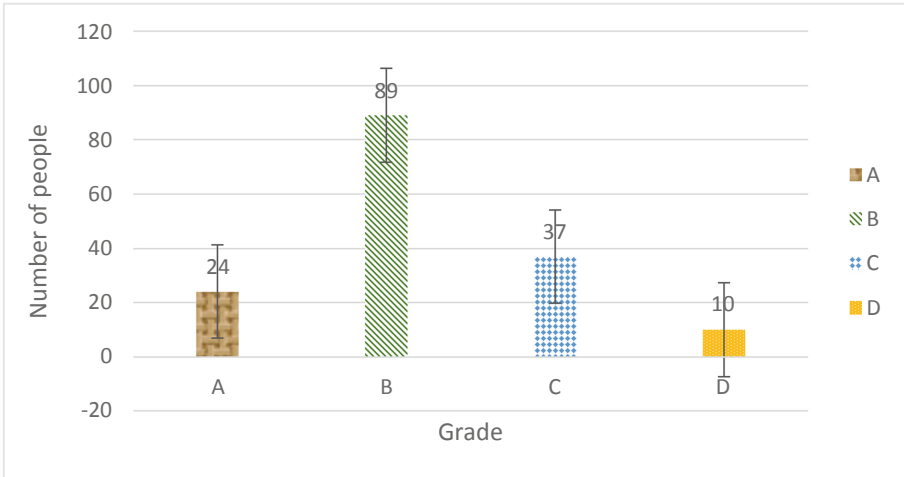


Fig. 3. Examination results statistics chart

When the test system feedbacks the test scores, 160 students are randomly selected to use the test system to take an oral test, with a total score of 50 points. The final results of oral English scores are divided into four levels: A, B, C, and D. The A-level test score corresponds to 40–50 points, the B-level corresponds to 30–40 points, and the C-level corresponds to 15–30 points. D level corresponds to 15 points or less. After the exam, the system will automatically count the number of people corresponding to each level, and

generate corresponding statistical charts, as shown in Table 2 and Fig. 3. From the data in the table, it can be seen that there are 24 people with oral test scores of A, accounting for 15% of the total, 89 people with B level, accounting for 55.625% of the total, and 37 people with C level, accounting for the total. There are 23.125% of the total number, and there are 10 people in the D-level, accounting for 6.25% of the total number. The use of this system for statistics of examination results saves teachers' time for statistical work and improves the efficiency of teacher assessment management.

5 Conclusion

The spoken English test system based on data mining algorithm designed in this paper is developed for the shortcomings of spoken English test. Through preliminary input into actual application operation, it is tested that the system can basically meet the requirements of spoken English test, which is improved to a large extent. The efficiency of the oral exam reduces the difficulty of organizing the exam and the intensity of the teacher's work.

References

1. Chai, Y.: Design and implementation of english intelligent communication platform based on similarity algorithm. *Complexity* **2021**(2), 1–10 (2021)
2. Denicola, N., Grossman, D., Marko, K., et al.: Telehealth interventions to improve obstetric and gynecologic health outcomes: a systematic review. *Obstet. Gynecol.* **135**(2), 371–382 (2020)
3. Yun, U., Yu, E., et al.: Design and Implementation of a friendly maze program for early childhood based on a path searching algorithm. *J. Korea Soc. Comput. Inf.* **22**(6), 49–55 (2017)
4. Shim, J.-S., Ju, Y.-W., Park, S.-C.: Design and implementation of a Pedestrian recognition algorithm using trilinear interpolation based on HOG-UDP. *J. Supercomput.* **74**(2), 787–800 (2017). <https://doi.org/10.1007/s11227-017-2160-1>
5. Nie, J.: Design and implementation of a data mining platform based on cloud computing. *Agro Food Ind. Hi Tech* **28**(1), 318–321 (2017)
6. Huang, W., Xiao, X., Xu, M.: Design and implementation of domain-specific cognitive system based on question similarity algorithm. *Cogn. Syst. Res.* **57**, 20–24 (2019)
7. Lewney, J.: Quality measures for dental care: a systematic review. *Evid. Based Dent.* **20**(3), 79–80 (2019)
8. Dugartysrenova, V.A., Sardegna, V.G.: Raising intercultural awareness through voice-based telecollaboration: perceptions, uses, and recommendations. *Innov. Lang. Learn. Teach.* **13**(3), 205–220 (2019)
9. Zhou, Y.: Design and implementation of book recommendation management system based on improved Apriori algorithm. *Intell. Inf. Manag.* **12**(3), 75–87 (2020)
10. Kalaiyarasi, D., Reddy, T.K.: Design and implementation of LMS adaptive FIR filter based on OBC distributed arithmetic algorithm with MCSLA. *J. Eng. Appl. Sci.* **14**(20), 7756–7764 (2019)
11. Nie, A.: Design of English interactive teaching system based on association rules algorithm. *Secur. Commun. Netw.* **2021**(s1), 1–10 (2021)
12. Song, S.: Design and implementation of speed detection and control module based on single-chip microcomputer control. *Revista de la Facultad de Ingenieria* **32**(14), 461–466 (2017)



Design of Intelligent Recognition System of Employees' Behavior in Coal Mine Based on VR Technology

Zenghui Liu^{1,2}(✉)

¹ China Coal Technology and Engineering Group Shenyang Research Institute, Fushun 113000, China

² State Key Laboratory of Coal Mine Safety Technology, Fushun 113122, China

Abstract. The video surveillance systems currently used for mine safety management are basically passive surveillance. The intelligent recognition technology of employee behavior in coal mines can analyze the video structure in real time and understand the image. Through the intelligent video recognition system, it can actively discover potential hazards, suspicious targets and abnormal behaviors, solve the lack of manual marking, give timely alarms and early warnings, and take timely actions for the administrator Prevent accidents with on-site staff. This article focuses on the research on the intelligent recognition system of underground coal mine employees' behavior based on VR technology. Based on the literature, it has a general understanding of the relevant technologies of the underground coal mine employee behavior intelligent recognition system, and then designs the underground coal mine employee behavior intelligent recognition system, and test the designed system, and the test results show that the Hu moment algorithm used in the system designed in this paper has good rotation, flip invariance, and high accuracy of comparison of the similarity of the pictures.

Keywords: VR technology · Intelligent recognition · Abnormal behavior · Coal industry

1 Introduction

Due to the frequent occurrence of coal mining accidents, the coal industry has become a high-risk industry, and the development of the coal industry is considered a long-term and arduous task [1, 2]. In addition, some of our research techniques have been applied to coal mining to monitor the condition of moving objects. Early common tracking technologies include radio frequency technology, tracking circuit detectors and infrared temperature measurement technology. With the development of computer networks and computer vision, the application of intelligent video surveillance technology emerged at the historic moment and was gradually applied to the field of coal mine safety [3, 4]. Traditional video surveillance systems usually record camera output. If there is an abnormality, the monitoring personnel can only know what happened from the recorded results, but it is often too late. The monitoring system based on intelligent video monitoring technology

can continuously monitor in real time, 24 h a day, and automatically analyze the image data recorded by the camera [5, 6]. In the event of a dangerous situation, the system can accurately and timely alert the monitoring personnel to avoid accidents in the dangerous areas of the mine. Therefore, the application of intelligent video monitoring technology in the underground environment of coal mines is of great significance for the safety of underground personnel and the efficient production of the mining industry [7, 8].

For the research of behavioral intelligent recognition, the current understanding of human behavior includes the definition and classification of human surveillance, formal gesture recognition, simple action recognition, and the most commonly used set of human actions (running, crouching, standing, jumping, climbing) [9]. Although in recent years, people use machine learning tools to establish the statistical model of human action awareness has achieved certain development, but human action awareness is still in its infancy. In the standard process of continuously matching ambiguity and reducing ambiguity, people often cause simplified constraints on human motion patterns, and this restriction is often consistent with the general image state [10]. Therefore, understanding its working mechanism has always been a difficult problem in feature selection theory and machine learning. At present, the state site method and pattern matching method of action recognition usually calculate the cost and accuracy of action recognition, but there is a trade-off between the two. Therefore, in order to improve the characteristics of action recognition, it is necessary to seek and develop new technologies to effectively reduce computational complexity [11]. In addition, more cutting-edge visual computing technology and artificial intelligence results can be used to extend the existing simple action recognition and semantic recognition to the natural language description of more complex scenes, and to promote the low-level computer vision processing capabilities [12]. In summary, there are more researches on intelligent recognition of human behavior, but less research on the application of behavior recognition to coal mine safety detection.

This article explores the intelligent recognition system for employee behavior in coal mines based on VR technology. Based on the literature, it analyzes the description of human behavior and the problems in the intelligent recognition system for employee behavior in coal mines, and then analyzes the behavior of employees in coal mines based on VR technology. The intelligent recognition system level is designed, and then the system is tested by experiments, and relevant conclusions can be drawn through the test results.

2 Research on Intelligent Recognition System of Employees' Behavior in Coal Mines

2.1 Description of Human Behavior

- (1) The contour feature is usually obtained by separating the foreground and background of the image to obtain the foreground boundary, and then using the edge detection operator to transform. Because contour features are not easily affected by color and texture, they are improved and adopted by many researchers. Contour feature extraction is convenient and not easily affected by color and texture. However, the human body's own occlusion has a greater impact on the contour feature extraction effect, and the occlusion problem has not been completely resolved.

(2) Optical flow characteristics

The visual flow function is a very good motion function, which is widely used in the field of human behavior recognition. The optical flow feature uses the optical flow detection method to detect the position and direction of the moving target, and the detection and recognition are carried out randomly through the sampling method, and the area of interest in the field of view is divided into blocks, and the optical flow is calculated according to the frame of the information box, and then add the optical flow information table of the same behavior, average the value to get the representative behavior, and finally use a simple classifier to identify the behavior.

(3) Temporal and spatial characteristics

Spatio-temporal features are defined as the local features of the 3D volume on the spatio-temporal scale, extracted from a small part of the local image sequence. The location and type of spatiotemporal features determine the representation of the video motion sequence.

(4) Depth characteristics

Compared with grayscale images, depth images are less affected by changes in illumination, and each pixel value represents the distance between the camera and the scene. The depth feature is widely used because it provides the 3D structure of the scene and can facilitate background removal, segmentation, and motion estimation.

In the process of recognizing human behavior, most researchers often choose individual characteristics to recognize behavior. Although this can simplify the identification process, there are usually problems with individual features that cannot be ignored. For example, the visual flow feature has a large amount of calculation, the contour feature has self-separation, and the spatio-temporal feature requires high preprocessing. These problems greatly affect the accuracy of behavior recognition. Therefore, in recent years, many domestic and foreign researchers have generally adopted multi-feature fusion strategies to obtain strong features in most studies. In other words, it analyzes many feature vectors (visual flow, color, texture, information depth, etc.). This performs simple combinations to form fusion properties that describe behavior. The combination function can more accurately characterize actions, greatly improving the accuracy and efficiency of behavior recognition. Therefore, multifunctional fusion technology is widely used in the field of human behavior recognition.

2.2 Problems in the Intelligent Recognition System for Employee Behavior in Coal Mines

- (1) Underground network structure is not simple, underground wiring is complicated, and maintenance is inconvenient. Software, hardware, and system wiring must run the “three networks in one” (including staffing system, monitoring and control system, communication and communication system) to help maintain the system.
- (2) The overall digital structure layout of the mine was not considered. This system is a very important part of mine IT work. It usually adopts a multi-layer network structure, including information collection level, data management level and network service level. In the system design scheme, if the system design is based on

the overall digital design of the mine and the future informatization requirements and the digital development of the mine, the rationality, economy, practicability and scalability of the system design will be stronger.

- (3) There is no systematic investigation of the height of the entire industry, and the technical standards, database standards, and interface specifications adopted for the integration of various systems are inconsistent, which is not conducive to the networking of the system.

2.3 Application of VR Technology

VR technology is a computer simulation of real-world scenes. In the intelligent system that recognizes coal miners' behavior, it is necessary to use VR technology to model human behavior and compare it with the collected data for identification.

3 Design of an Intelligent Recognition System for Employees' Behavior in Coal Mines Based on VR Technology

3.1 System Design Analysis

After successfully detecting a target moving within the monitoring range, it is necessary to determine whether the target is operating abnormally and issue an alarm.

Before the discovery of abnormal behavior, a clear supervision of legal behavior is required to establish a set of rules. This paper uses the method of creating a database to store the moving parts of the target object, and then runs the template mapping query in the current frame for anomaly detection.

The specific implementation process is as follows.

- (1) Before watching the video, first save a series of regular moving images in the database. Due to the symmetry of the images in the library, they are divided into two categories, from left to right and from right to left, depending on how you walk. This is called the "normal motion pattern library".
- (2) In the real-time monitoring stage, the video stream is read frame by frame, and each frame of image is grayed out and morphologically processed. The database template library contains the corresponding category, and the current frame is all images compare to. After the comparison, a set of error values $I(A, B)$ are obtained.
- (3) Compare all error values in the result, take the minimum value and compare with the default threshold. If it is within the specified threshold, it is a legal behavior, otherwise it is an abnormal behavior, and follow-up alarm processing is performed.

3.2 Acquisition Module

The collection terminal adopts explosion-proof design, the explosion-proof grade is ExdiblIB + H2T4Gb, and the protection grade is IP65. Module storage space adopts independent design, including intrinsically safe circuit cavity, explosion-proof cavity and gas sensitive cavity. The device incorporates technologies such as video surveillance, gas detection, and multiple wireless transmission GPS/Hokuto detection (supporting

3G/4G and Huawei private network). Multiple sensors integrate cloud functions such as explosion protection, site geographic display, gas gathering alarm, local and remote control. Among them, the gimbal is made of high-strength aviation aluminum, which can rotate 360° infinitely. The battery is a large-capacity low-temperature lithium battery, equipped with a battery certified by a national professional organization. The life span is more than 11 h and the working environment is $-259\text{ }^{\circ}\text{C}$ – $+55\text{ }^{\circ}\text{C}$. The camera is a high-resolution digital camera with adjustable focal length, which can adapt to the requirements of camera accuracy in various working environments.

3.3 Identification Module

(1) Feature extraction

In the physical sense, kinetic energy represents the energy generated by the movement of an object, and the average kinetic energy of the group represents the intensity of the group's movement. A group can be considered as a collection of many atoms, and these atoms are linearly related to the AGAST corner points in the binary foreground mask area, so the AGAST corner points can be used to represent the moving atoms. Therefore, the calculation of the group average kinetic energy uses the AGAST corner points as the calculation feature points based on the flow field of view. In other words, the total kinetic energy of the group can be expressed by calculating the luminous flux energy of the AGAST corner points in the range of the binary foreground mask. Therefore, the calculation of the group average kinetic energy is also the calculation of the average angular kinetic energy. According to the definition of kinetic energy of moving objects in physics, the average kinetic energy of group motion in each frame can be expressed by Eq. (1).

$$E_k = \frac{\sum_{i=1}^n m_i |v_i|^2}{n} \quad (1)$$

Among them, E_k represents the average kinetic energy of the group, and n represents the total number of moving targets, that is, the number of characteristic corner points. m_i represents the qualitative weight value of the motion vector i . Assuming that the individuals in the group are similar in size and weight, then m_i is a constant.

The location distribution among individuals in a cluster is the main indicator of the characteristics of group behavior. However, because of the large number of individuals in the cluster, it is impossible to visually describe the relative position distribution of individuals in the location distribution, so it is used to indirectly explain the relative status of individuals in the entire cluster. Therefore, by estimating the average of the coordinates of each person's tilt point, this paper can approximate the average value of the center of the region where the group is located, and atomize the atom with the euclidean distance from the corner point to the center of the group. Use Eq. (2) to calculate $D(x)$ in the x -th frame.

$$D(x) = \varphi \frac{\sum_{i=1}^N \sum_{j=i+1}^N C_{ih}}{(N+1)!} \quad (2)$$

Among them, C_{ij} is the Euclidean distance between two points, φ is a correction coefficient considered as a constant, and N is the total number of AGAST corner points in the field mask.

The feature extraction can be divided into the following points. The foreground image is received by performing $5 * 5$ kernel expansion on the binary image, and then the field information is received by the optical flow method and the average value is extracted. Kinetic energy and directional kinetic energy entropy group dynamics characteristics, $3 * 3$ nuclear corrosion is performed behind the nuclear expansion image, and then the euclidean distance and distance potential energy between the AGAST corner points in the field of view mask are calculated.

(2) Identification process

1) Identification.

In the above abnormal behavior detection process, the most critical step is to compare the current frame with the standard image stored in the library, that is, how to calculate the similarity detection result of the two images to the greatest extent. After repeated tests and comparisons, this article uses this model as a reference.

(2) Classification

The classification in this paper is based on neural networks. The structure of a typical single-hidden neural network power supply layer is shown in Fig. 1. The plane has a neuron, and for each input variable, the hidden layer has a neuron.

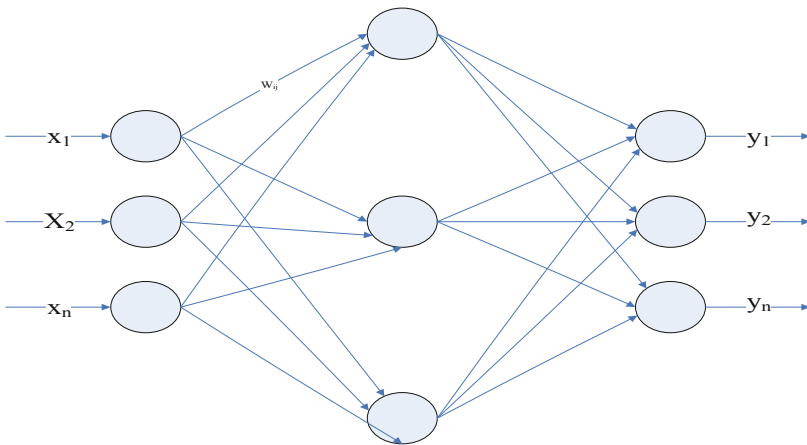


Fig. 1. Network structure of extreme learning machine

4 System Test

The Hu torque algorithm of this paper is tested here to test the theory of unchanged Hu torque. Table 1 shows the data obtained by exporting the Hu torque characteristics to binary images of six aircraft.

Table 1. System test results

	M1	M2	M3	M4	M5	M6	M7
1	3.81	5.41	7.61	6.21	13.09	8.89	12.55
2	3.82	5.62	7.68	6.32	13.13	9.95	13.28
3	3.80	5.64	7.60	6.23	12.89	8.93	12.72
4	3.71	6.32	6.41	4.33	9.31	7.34	9.65
5	3.72	6.33	6.40	3.42	8.13	4.78	8.18
6	3.37	6.35	6.53	3.97	8.12	4.75	8.52

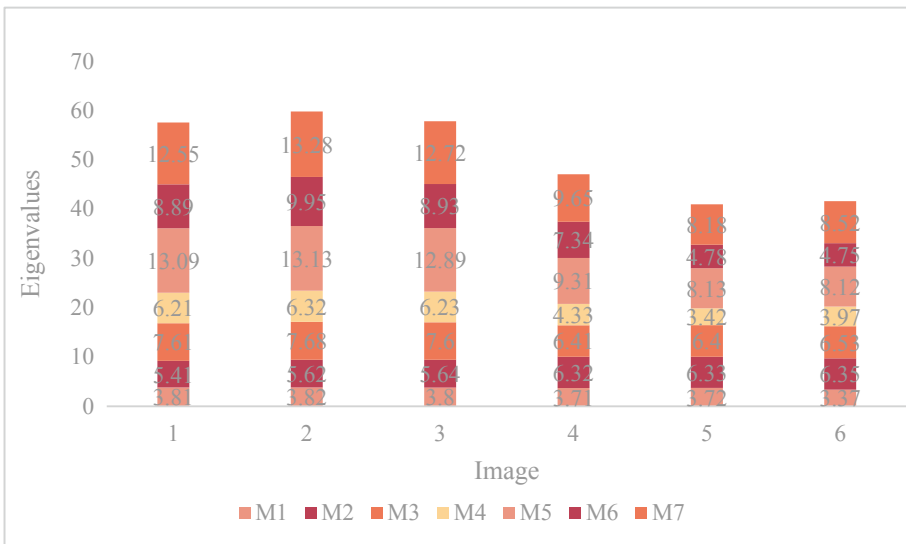


Fig. 2. System test results

It can be found from Fig. 2 that the eigenvalue results of the grade variables of the three images 1, 2, and 3 are relatively similar, while the eigenvalue results of the grade variables of the three images 4, 5, and 6 are relatively similar. In fact, when sorting using a specific grouping algorithm, it can often be found that the first three images are in the same type, and the last three images are in the same category. This also proves that the Hu torque algorithm can perform good rotation without adjustment, and the similarity of the graph is inverted, which is more accurate.

5 Conclusions

This article focuses on the research of the intelligent recognition system of coal mine employees' behavior based on VR technology. After understanding the relevant theories, the design of the intelligent recognition system of coal mine employees' behavior based

on VR technology is carried out, and the designed system is tested. The test results It is concluded that the Hu torque algorithm used in this paper can achieve good rotation without modification, image similarity inversion, and relatively high accuracy.

References

1. Wang, Z., Guo, B., Yu, Z., et al.: Wi-Fi CSI based behavior recognition: from signals, actions to activities. *IEEE Commun. Mag.* **56**(5), 109–115 (2018)
2. Helbing D, G Tröster, Wirz M, et al.: Recognition of crowd behavior from mobile sensors with pattern analysis and graph clustering methods. *Netw. Heterogen. Media* **6**(3), 521–544 (2017)
3. Yousefi, S., Narui, H., et al.: A survey on behavior recognition using WiFi channel state information. *IEEE Commun. Mag.* **55**(10), 98–104 (2017)
4. Guo, B., Chen, Y.J., Lane, N., et al.: Behavior recognition based on Wi-Fi CSI: Part 1. *IEEE Commun. Mag.* **56**(5), 108 (2018)
5. Yang L, Ma R, Zhang H M, et al.: Driving behavior recognition using EEG data from a simulated car-following experiment. *Accid Anal Prev*, 2017, 116(JUL.):30–40
6. Chen, L., Chen, X., Ni, L., et al.: Human behavior recognition using Wi-Fi CSI: challenges and opportunities. *IEEE Commun. Mag.* **55**(10), 112–117 (2017)
7. Guo, Y., Meng, W., Fan, Y., et al.: Wearable sensor data based human behavior recognition: a method of data feature extraction. *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/J. Comput. Aided Des. Comput. Graph.* **33**(8), 1246–1253 (2021)
8. He, Y., Gong, X., Yuan, C., et al.: Lateral obstacle avoidance control based on driving behavior recognition of the preceding vehicles in adjacent lanes. *Automatisierungstechnik* **68**(10), 880–892 (2020)
9. Pang, C.: Simulation of student classroom behavior recognition based on cluster analysis and random forest algorithm. *J. Intell. Fuzzy Syst.* **40**(2), 2421–2431 (2021)
10. Zheng, B., Yun, D., Liang, Y.: Research on behavior recognition based on feature fusion of automatic coder and recurrent neural network. *J. Intell. Fuzzy Syst.* **39**(6), 8927–8935 (2020)
11. Qu, J., Qiao, N., Shi, H., et al.: Convolutional neural network for human behavior recognition based on smart bracelet. *J. Intell. Fuzzy Syst.* **38**(8), 1–12 (2020)
12. Wu, S.: Simulation of classroom student behavior recognition based on PSO-kNN algorithm and emotional image processing. *J. Intell. Fuzzy Syst.* **40**(4), 1–11 (2020)



Design of Chinese Linguistics Teaching System Based on K-means Clustering Algorithm

Hongling Ma^(✉)

Inner Mongolia Vocational College of Chemical Engineering, Hohhot, China

Abstract. With the rapid development of Internet technology, a variety of network application software have sprung up. The traditional teaching model can not meet the learning needs of students. The education websites that have sprung up and grown rapidly under the web environment just make up for this defect. Therefore, based on K-means clustering algorithm, this paper designs and develops a Chinese linguistics course system. Firstly, this paper expounds the concept of MOOC system, and then explains the development process of Chinese Linguistics network teaching. Then, based on K-means clustering algorithm, this paper designs and develops the framework of Chinese Linguistics MOOC system, and analyzes the audience, student group and system performance of Chinese Linguistics MOOC system respectively. Finally, the test results show that, the response time and delay time of each module of the system are basically about 2–3 s, and the utilization rate of the client CPU is also about 1%–2%, and the operation process shows that the performance of the system can meet the use needs of users.

Keywords: K-means clustering algorithm · Chinese Linguistics · MOOC class system · System design

1 Introduction

Chinese Linguistics Mu course system is a rich network learning resources, teaching information and service functions, which provides educators with efficient, comprehensive, safe and reliable online browsing of various courses [1, 2]. With the development of Internet technology, network has become an indispensable part of people's life and learning. In the rapid development of computer, it is urgent to develop a new efficient, safe and reliable online communication environment based on database system [3, 4].

Many scholars have studied the Chinese language learning system. Some scholars have proposed a new data model, which can extract, classify and map data to generate more structured new data to meet organizational needs. This can be achieved by calculating the weight of various metadata attributes [5, 6]. Some scholars have proposed an interactive language teaching system based on association rule algorithm. This method can make full use of the correlation between data and improve the efficiency of language teaching. According to the division of UNESCO, educational informatization is divided into four stages: start-up, application, integration and innovation. With the initial development of network hardware technology construction in recent years, it has gradually

entered the integration and innovation stage of education and information technology [7, 8]. The above research has laid a certain foundation for the design and research of Chinese Linguistics MOOC system based on K-means clustering algorithm.

This paper presents a new function of Chinese Linguistics MOOC system based on K-means clustering algorithm. This design puts forward and implements the information interaction function module according to the analysis and research of the Chinese language learning background, so as to improve the teaching quality and efficiency of school MOOC, and uses K-means clustering technology to make a complete organic framework architecture for the data structure to support the operation of the whole platform, including user login, registration management and teacher registration.

2 Discussion on the Design of Chinese Linguistics MOOC System Based on K-means Clustering Algorithm

2.1 MOOC System

MOOC course has changed the traditional teaching method, and the content of MOOC course is also very different from traditional courses [9, 10]. After a comprehensive analysis of the MOOC courses on the most important MOOC platforms such as udacity, coursera and EDX, it can be concluded that most of the MOOC on the MOOC platform are composed of the following content modules:

(1) Course content

Weekly course video is the most important knowledge transfer in MOOC, and it is also the most important part of MOOC. MOOC mainly focuses on high-quality short video. The video duration is between 5 and 10 min. Some videos also come with ppts to make them more visual, and some provide additional reading materials. There are also Q & A videos to keep students' attention and test their understanding. Only by correctly answering relevant questions can they watch the following videos [11, 12].

(2) Community chat

Because there are too many learners in MOOC, it is impossible for teachers to answer each student's questions. Therefore, learners of the same course form a specific learning community on the Internet and use this platform for learning and communication. On this platform, learners can not only ask questions, but also view the communication content of other learners and benefit from it. Sometimes students can learn what is not covered in MOOC courses through the platform. Of course, some teachers will use group chat to answer students' questions during "online consultation time".

(3) Curriculum evaluation

In addition to embedding tests in instructional videos, MOOC teachers can also create assignments and conduct tests. "The well-designed and thoughtful MOOC platform provides comprehensive job testing functions, such as traditional teaching. Due to the large number of course participants, teachers and teaching assistants cannot evaluate and process them alone. Therefore, evaluate homework or MOOC - mainly through computer evaluation or peer review.

2.2 Development of Chinese Linguistics Course

MOOC model is the product of the development of online courses so far. It has experienced three stages of development: the first is the digital course represented by MIT, the second is the prestigious open video course, and the third is MOOC. With the increasing number of Chinese learners and the popularity of the Internet, the Chinese Teaching community has also begun to use online courses for teaching. This is convenient and efficient. The free learning method not only breaks the time and space of the traditional Chinese classroom, but also can use modern technology to promote Chinese to the world. At present, MOOC teaching is in the stage of rapid development. Different disciplines are trying to combine with MOOC teaching, and different MOOC courses have sprung up. However, there are few Chinese linguistics teaching in MOOC, and MOOC has entered the world. It is only two and a half years in China. Of course, the production technology has not been popularized. How to complete the interaction between teachers and learners has not been well clarified, which also affects the teaching effect of MOOC course to a certain extent. Therefore, the development of Chinese Linguistics still needs a lot of research and work.

2.3 K-means Clustering Algorithm

2.3.1 Clustering Idea

K-means clustering algorithm is one of the most basic and widely used classification methods in cluster analysis. It is a method to find clusters and cluster centers in labeled data without categories. Select the number of K cluster centers you want, and the K-means process will repeatedly move the centers to minimize the variance of the whole cluster. The basic idea of the algorithm is as follows: a given cluster has n clusters. The database of the data object and the number of K clusters to be generated, randomly select K objects as the initial K cluster centers, then calculate the center of the remaining sample cluster for each cluster, place the sample in the center of the next cluster, and calculate the new cluster center with the mean method for the newly fitted class. If there is no cluster center in the two adjacent clusters, any change indicates that the sample adjustment is complete and clustered. The criterion function e of the class average error has converged. During each iteration, the algorithm must check whether the classification of each sample is correct. If it is not correct, it needs to be adjusted. After adjusting all samples, change the cluster center and enter the next iteration. If all samples are correctly classified in the iterative algorithm, there is no adjustment or change in the cluster. During the iteration of the algorithm In the process, the value of E decreases continuously and finally converges to a fixed value. This standard is also one of the basis to measure whether the algorithm is correct.

2.3.2 Algorithm

In the clustering algorithm, there are many types of calculation functions for the distance between data objects, which are mainly divided into absolute distance, Minkowski distance (including absolute distance, Euclidean distance and Chebyshev distance), Mahalanobis distance and Canberra distance. Among most clustering algorithm applications,

Euclidean distance method in nominal distance is a more common method to calculate the distance between numerical objects. The specific contents include:

$$d_{ij}(q) = \left(\sum_{k=1}^q x_{ik} - x_{jk} \right)^q, q > 0 \quad (1)$$

$$d_{ij}(1) = \left(\sum_{k=1}^q x_{ik} - x_{jk} \right)^q, q = 1 \quad (2)$$

$$d_{ij}(2) = \left(\sum_{k=1}^q x_{ik} - x_{jk} \right)^q, q = 2 \quad (3)$$

In addition, the most common methods to calculate the distance between data objects in clustering algorithm include chi square distance and oblique space distance. Because the clustering algorithm can be applied to all aspects of life, the clustering results of the algorithm are obtained by calculating the distance between different data objects in different fields, so there are some differences. Therefore, in practical algorithm application, we often use a variety of distance calculation methods for distance calculation. Finally, we compare and analyze the grouping results, and finally calculate the distance calculation method with better aggregation result quality.

3 Experiment

3.1 Development Basis of MOOC System

Online education is a beneficiary of the education and teaching sector. Computers and the Internet are used for distance learning. It is the inevitable outcome of the development of information technology and education. It is an important means of distance learning and lifelong learning. It provides an essential way for students to obtain high-quality educational resources. In online teaching, tutorials and teaching materials (DOC, PPT, PDF, txt) are currently being uploaded to the online teaching platform at a specified time, and online courses open teaching tutorials for students to study independently. This method not only trains students to acquire knowledge and information, but also cultivates students' self-study ability and the ability to acquire knowledge by using the network. It can significantly expand the target group of the class and improve students' learning ability. The scale of school management can reduce teaching costs. This kind of education is characterized by educational informatization, diversification of forms, students' subjectivity and school openness. MOOC system mainly provides online teaching materials and tutorials, online simulation exams, online Q&A and discussion rooms, online teaching videos and videos. The characteristic of the system is that it is always designed to be interactive. Considering the openness of the education system, it is usually designed to be modular and scalable. In terms of technical implementation, B/S mode is generally adopted, and the development language should be ASP, SP and PHP.

3.2 Framework of MOOC System

Figure 1 is the frame structure of Chinese linguistics teaching system. The student function subsystem provides services for students' Online Chinese Linguistics courses, which is the main functional component of the system, including many specific learning functions, such as Chinese language course learning, after-school homework, course examination, course forum, Chinese Linguistics data learning, student user personal data, etc. The teacher function subsystem provides support for teachers to teach Chinese Linguistics courses online. It is an important functional part of the system, including course content management, examination management, question answering and so on. The system management module is a necessary part of the system and is indispensable. It is mainly responsible for user and authority management, website related settings, etc.

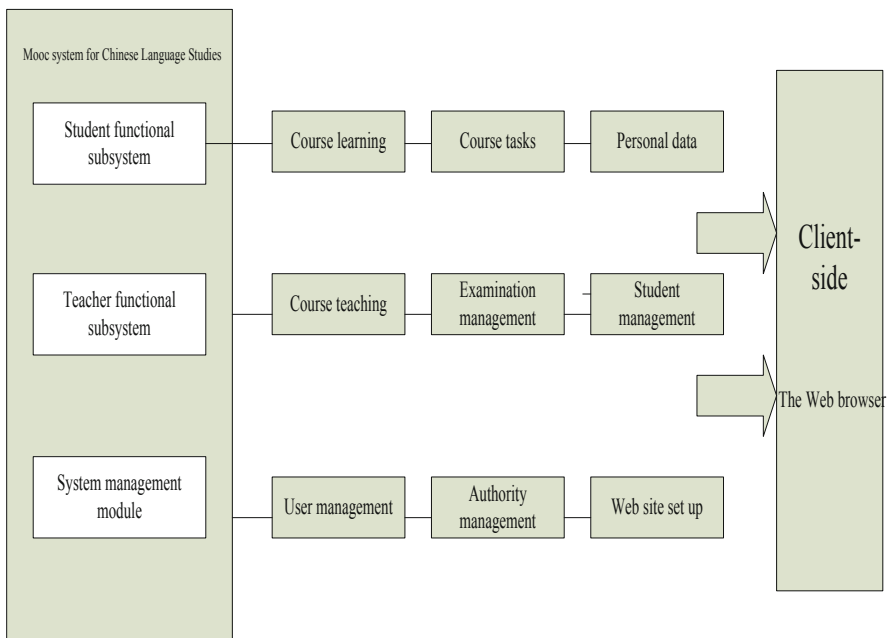


Fig. 1. Framework diagram of the Chinese Language MOOC teaching system

4 Discussion

4.1 Content Analysis of Chinese Linguistics Course for Learners

Table 1 shows the needs of learners of Chinese Linguistics in MOOC.

As can be seen from Table 1, the learners of Chinese Linguistics may include two categories. The first is systematic Chinese learners. They may have learned Chinese, but they can have smooth expression ability. Therefore, they choose this course to consolidate

the learned content and lay a good foundation for future learning. The second is non systematic Chinese learners, mainly foreigners who want to learn basic Chinese. Most of them have never studied Chinese. They are zero starting point learners. Their purpose of learning this course may be to travel to China, business communication or just because they are interested in Chinese. Therefore, they choose this course just to master and understand some basic Chinese knowledge and basic daily communication language.

Table 1. Analysis of Chinese language course learners

Course title	Chinese linguistics		
The learner analysis	Types of learners	Learning needs	The first foundation of learners
	Systelearning systematically	Professional Chinese knowledge requirements; language skills and speech communication skills requirements	Learners are not required to master, any Chinese language knowledge first
	Non-systematic Chinese-language learners	Daily communication of tourism and business; interested in Chinese and Chinese culture	Learners are not required to master, any Chinese language knowledge first

4.2 Performance Analysis of Chinese Language Teaching System

Table 2. Performance analysis of Moa System

System module	Delay time (s)	CPU occupancy rate (%)	Reaction time (s)	Is the operation smooth
Student functional subsystem	1	2	2	Yes
Teacher function subsystem	3	2	3	Yes
System Management Module	2	1	3	Yes

This paper tests the performance of each module of the Chinese linguistics teaching system, which is the student function subsystem, the teacher function subsystem and the system management module. It can be seen from Table 2 and Fig. 2 that the response time and delay time of each module of the system are basically about 2–3 s, and the utilization rate of the client CPU is also about 1%–2%, and the operation process shows that the performance of the system can meet the use needs of users.

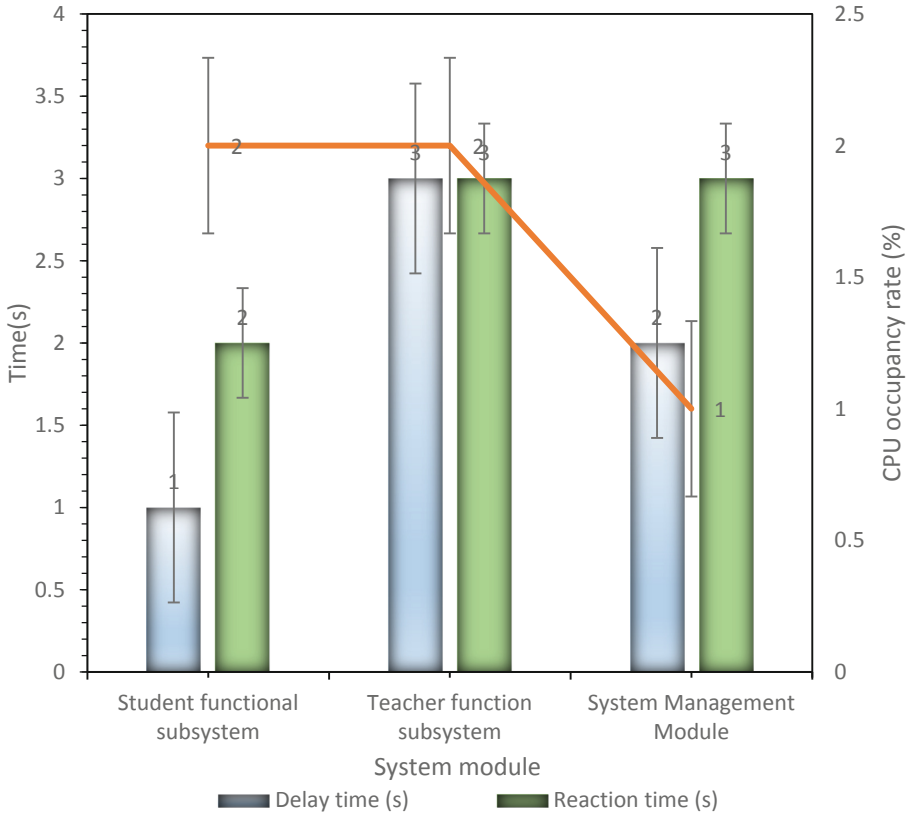


Fig. 2. System performance test

5 Conclusion

With the development of network technology, new changes have taken place in the field of education. Multimedia teaching has been gradually popularized. How to use advanced information technology to improve school management and service level is an urgent problem to be solved. Chinese linguistics course system is a new and efficient learning website. This paper mainly introduces the application status and existing problems of K-means clustering algorithm (WEB middle school teachers and students register as members), analyzes and studies it, and designs a set of teaching auxiliary tools in the network environment with reasonable requirements, powerful functions and high practicability, hoping to help the majority of geography teachers improve teaching efficiency and service quality.

References

1. Kachaoui, J., Belangour, A.: Enhanced data lake clustering design based on K-means algorithm. *Int. J. Adv. Comput. Sci. Appl.* **11**(4), 547–554 (2020)

2. Kandali, K., Bennis, L., Bennis, H.: A new hybrid routing protocol using a modified K-means clustering algorithm and continuous hopfield network for VANET. *IEEE Access* **9**, 47169–47183 (2021)
3. Shrifan, N.H.M.M., Jawad, G.N., Isa, N.A.M., Akbar, M.F.: Microwave nondestructive testing for defect detection in composites based on K-means clustering algorithm. *IEEE Access* **9**, 4820–4828 (2021)
4. Sukanya, S.T., Jerine: A novel melanoma detection model: adapted K-means clustering-based segmentation process. *Bio Algorithms Med Syst.* **17**(2), 103–118 (2021)
5. Wei, L.X.: The analysis and implementation of the K - means algorithm based on Hadoop platform. *Comput. Inf. Sci.* **11**(1), 98–103 (2018)
6. Chen, D., Jin, C., Cao, Y., et al.: Segmentation and classification of brain anatomy in magnetic resonance images based on fuzzy C-means clustering algorithm. *J. Med. Imaging Health Inf.* **7**(4), 844–849 (2017)
7. Siamala Devi, S., Anto, S., Siddique Ibrahim, S.P.: An efficient document clustering using hybridised harmony search K-means algorithm with multi-view point. *Int. J. Cloud Comput.* **10**(1/2), 129–143 (2021)
8. Zhu, B., Bedeer, E., Nguyen, H.H., Barton, R., Henry, J.: Improved soft-k-means clustering algorithm for balancing energy consumption in wireless sensor networks. *IEEE Internet Things J.* **8**(6), 4868–4881 (2021)
9. Liu, J., Yang, B.: Design of evaluation system of physical education based on machine learning algorithm and SVM. *J. Intell. Fuzzy Syst.* **1–3**, 1–12 (2020)
10. Daoudi, S., Zouaoui, C.M.A., El-Mezouar, M.C., Taleb, N.: Parallelization of the K-Means++ clustering algorithm. *Ingénierie des Systèmes d Inf.* **26**(1), 59–66 (2021)
11. Kumar, D.M., Satyanarayana, D., Prasad, M.N.G.: MRI brain tumor detection using optimal possibilistic fuzzy C-means clustering algorithm and adaptive k-nearest neighbor classifier. *J. Ambient. Intell. Humaniz. Comput.* **12**(2), 2867–2880 (2020). <https://doi.org/10.1007/s12652-020-02444-7>
12. Klopotek, M.A., Klopotek, R.A.: Issues in clustering algorithm consistency in fixed dimensional spaces. Some solutions for k-means. *J. Intell. Inf. Syst.* **57**(3), 509–530 (2021)



Construction of Maker Educational Resource Sharing Platform Based on Web Technology

Xia Li()

Zibo Normal Collage, Zibo 255000, Shandong, China

Abstract. With the advent of the era of knowledge economy, knowledge, as an intangible resource, is gradually replacing material resources in the traditional sense and becoming a key factor in promoting social progress. The openness and exchange of knowledge has become an urgent need for people. With the popularization of the global Internet, various resource management platforms based on network technology have been favored by all walks of life, and enterprises in the education field have become the focus of attention in particular. Maker training has unlimited potential and will have a significant impact on personal development, team curriculum reform, national education system and talent strategy reforms. This article is a research on the construction of a web technology-based maker education resource sharing platform. Its main purpose is to better publicize the importance of innovation, let more industry participants understand the concept of innovation and entrepreneurship and its industry characteristics, and promote the innovation and entrepreneurship movement development of. This article creates a platform for sharing educational resources and provides creators with an opportunity to showcase their creative works. Through the projects displayed on the platform and related tutorial videos shared by platform users, more people can learn about Makers. The final results of the study show that self-efficacy refers to having sufficient self-confidence in completing the work by themselves. 37.72% and 10.53% of learners respectively showed more self-confidence and self-confidence, so nearly half of the learners showed higher self-confidence.

Keywords: Web technology · Resource management · Maker education · Sharing platform

1 Introduction

With the popularity of the Internet around the world, various Web development technologies have become increasingly mature. However, due to the development of various programming languages and the emergence of various excellent and stable system frameworks, the continuous updating of technology has promoted the emergence of new products, which has given birth to a new industrial structure. In this era of rapid technological development, the diversification of products in various industries has given birth to new demands and the emergence of various new types of workers in society [1]. In the field of innovation, there has been a group of people who have a strong interest and hobby for new things. They try to prove their creativity through practice

and are committed to turning creativity into reality. People call it “Maker”. Maker is committed to innovation and construction, with vivid thinking, improve people’s living standards, improve the production efficiency of industrial products, and promote social and economic development.

In recent years, many researchers have conducted research on the construction of web technology-based maker education resource sharing platform, and have achieved good results. For example, Liu S believes that compared with traditional education, maker education incorporates innovative thinking and innovative spirit. More emphasis and emphasis on experiential education, with a better experience model in the interactive experience and sympathetic experience of teachers and students [2]. Wang P believes that Maker Education emphasizes equal participation in the participation model, not limited to subjects, ages and grades. Maker education teachers fully display the students’ personality through the regulation and management of the learning situation, not just limited to the knowledge in the books [3]. At present, scholars at home and abroad have conducted a lot of research on the sharing platform of maker education resources on web technology. These previous theories and experimental results provide a reliable theoretical basis for the research content of this article.

This article is based on web technology to realize the construction of a maker education sharing platform. Knowledge exchange is indispensable in the process of maker education, because knowledge can only realize its value in communication, and only communication can bring about knowledge innovation, so creation The demand for knowledge exchange will inevitably continue to grow. Open and shared knowledge has become an urgent need for people. Based on existing research, this article can establish that the learning process of educational university builders is from the traditional teaching centered on subject knowledge and reasoning as the main body. And open resource sharing is an important platform for makers to innovate and improve their practical ability.

2 Research on the Application of the Maker Education Resource Sharing Platform

2.1 Related Technologies of Maker Education Sharing Platform

(1) Spring framework

Spring is a lightweight framework designed to reduce the complexity of the software design process. Developers can use the basic Bean classes in the framework to implement many complex business requests. With the continuous update of technology, it has been widely used in transaction management, automatic dependency injection, and receiving exceptions. Generally speaking, the Spring framework has two key features: AOP and IoC. AOP stands for aspect-oriented programming. The main design idea is to isolate the design logic of the code from the corresponding service. It mainly obtains the internal information of objects in Java through “cross-cutting”, then derives their common characteristics, and integrates the content into a new part called “Aspect”. The main function of this unit is to make the pairing rate between the modules of the system lower, while reducing the code repetition

rate, making the operation and maintenance of the application system more convenient. IoC, that is, inversion of control. From the developer's point of view, IoC is an object-oriented programming design idea that can guide developers to design better applications. In the native system architecture, every time a program runs and uses a class object, the server must refresh the corresponding JavaBean for it [4, 5]. Once the framework adopts IoC, the container will automatically insert the JavaBean of each class object into the system according to the corresponding configuration information. This not only reduces the memory consumption on the server side, but also shortens the time to create each JavaBean object, thereby improving the processing speed of the business request system.

(2) Hibernate framework

Hibernate is a framework for modeling database tables. The framework realizes the one-to-one mapping relationship between the variables of the Java object class and the table fields in the database by encapsulating the native JDBC, making the module design logic clearer. Developers can also directly create corresponding SQL statements through this mapping relationship and the logical structure of the database, so that the pairing of databases and objects becomes very low. Compared with traditional SQL query statements and native JDBC drivers, the Hibernate framework optimizes the corresponding methods on this basis to make it more efficient. When the background module uses these methods to operate on the data in the linked database, the system memory consumption and the time required to retrieve data can be significantly reduced [6, 7]. The operation functions of Hibernate and Mybatis in the whole system framework are similar. But the difference between the two is that Hibernate can rely on its pre-designed mapping relationship to make Hibernate database highly portable. The Hibernate framework also provides a large number of log information interfaces, including program exception information, user logs and operation logs, system cache messages, and the display of SQL statements.

(3) PostgreSQL framework

PostgreSQL is an open source relational database system. The most basic requirement of each database system is to ensure the security, consistency and stability of the stored data [8]. After years of development and improvement, the PostgreSQL system can not only meet the requirements, but also can quickly store massive amounts of data. PostgreSQL can run stably on all major operating systems. It supports various database design patterns, creation of views, design of triggers, and development of stored procedures, etc. PostgreSQL has a high degree of scalability between massive data storage and high concurrent access, and can meet the development requirements of high concurrent application systems.

2.2 Elements of the Maker Educational Resource Sharing Platform

(1) Knowledge sharing subject

In the process of knowledge sharing in maker education, learners, teachers and maker spaces are the main body of knowledge sharing, and they are both the providers and receivers of knowledge.

(2) Knowledge sharing options

Knowledge sharing In knowledge management, knowledge service providers have evaluated, based on general factors, their intentions and platform incentives, and other individual knowledge sharing behaviors, etc., make the final decision whether to share knowledge with other learners. A decision may be different, it is a long-term, continuous choice [9, 10].

(3) Knowledge sharing process

Knowledge sharing refers to the process of voluntary mutual transfer and sharing of knowledge between different subjects of knowledge. The knowledge sharing in the knowledge sharing platform mainly involves two ways. The first is the transfer and exchange of knowledge between individuals, between individuals and groups, and between groups on the platform. For example, some members post. They display their experience in the project, their skills and knowledge in form, and other members receive knowledge through browsing and downloading [11, 12]. And combine their own relevant knowledge to put forward suggestions or opinions; second, the knowledge exchange platform itself provides learning knowledge with knowledge as the main body. Knowledge about the platform will also be provided. The specific performance is that the platform collects, organizes, records and records the knowledge information generated in the member communication activities on the platform. Members can also search for the knowledge they need on the platform.

3 Research on the Application Experiment Preparation of the Maker Education Resource Sharing Platform

3.1 Experimental Method

In the model of the maker education resource sharing platform, the semantic space vector model is usually used for research. The similarity between two different texts can be expressed by the cosine angle:

$$Sim(d, q) = dq = \sum_i (a_i \times b_i) \quad (1)$$

$$Sim(d, p) = \frac{dq}{||d|| \times ||q||} = \frac{\sum_i (a_i \times b_i)}{\sqrt{\sum_i a_i^2 \times b_i^2}} \quad (2)$$

Among them, a_i and b_i represent the frequency of each word in d and q , respectively. If the cosine value is zero, it means that the vector of the search term is perpendicular to the vector of the document, that is, there is no match, which means that the document does not contain the search term.

3.2 Experimental Data Collection

The experimental data in this paper is obtained through a questionnaire survey. The research objects include learners who have received maker education in colleges and universities, members of maker space, and entrepreneurs at school. The definition of

the research objects is mainly due to the knowledge of innovation and entrepreneurship of this type of learners., Can more accurately reflect the current development of maker education, and at the same time, this type of interviewees have experience in using knowledge sharing platforms in their usual learning process. A total of 200 questionnaires were distributed in this survey, and random surveys were conducted on the research subjects.

4 Analysis of the Experimental Results of the Application of the Maker Education Resource Sharing Platform

4.1 Statistical Analysis of Individual Factors

Learners are the main subject of knowledge sharing in the knowledge sharing platform. Whether they are willing to share knowledge with others, as well as the content and frequency of sharing are important factors affecting knowledge sharing. The experimental research in this article mainly focuses on self-efficacy, expectation of results, and evaluation scruples. The experimental results are shown in Table 1.

Table 1. Individual factor survey statistical analysis table

Options	Self-efficacy (%)	Result expectation (%)	Evaluation scruples (%)
Totally inconsistent	8.77	12.28	9.65
Less compliant	16.67	14.91	18.42
Uncertain	26.31	19.30	21.05
More in line with	37.72	36.84	36.84
Completely suitable	10.53	16.67	14.04

As shown in Fig. 1, by comparing the data, we can see that self-efficacy refers to having sufficient self-confidence in completing the work by themselves. 37.72% and 10.53% of the learners showed more self-confidence and self-confidence, so close to half Learners show higher self-efficacy. The result expectation means that the learner can achieve good expectations for this result. There are 36.84% and 16.67% who agree and agree very much, which shows that learners show higher expectation of results. Among the appraisal concerns, 36.84% and 14.04% felt that they were more concerned and anxious. This was because some students were afraid that the content shared would not be liked by others.

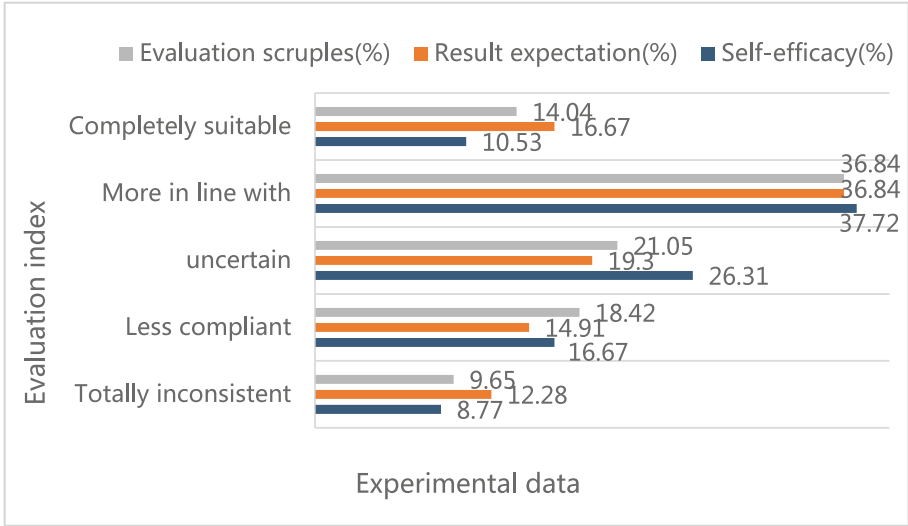


Fig. 1. Statistical analysis chart of individual factor survey

4.2 Statistical Analysis of Study Space Atmosphere Survey

The spatial atmosphere factors of the maker’s educational resource sharing platform in this paper mainly focus on the two dimensions of reciprocity and trust, and subdivide trust into emotional trust and economic trust. The experimental results are shown in the figure below.

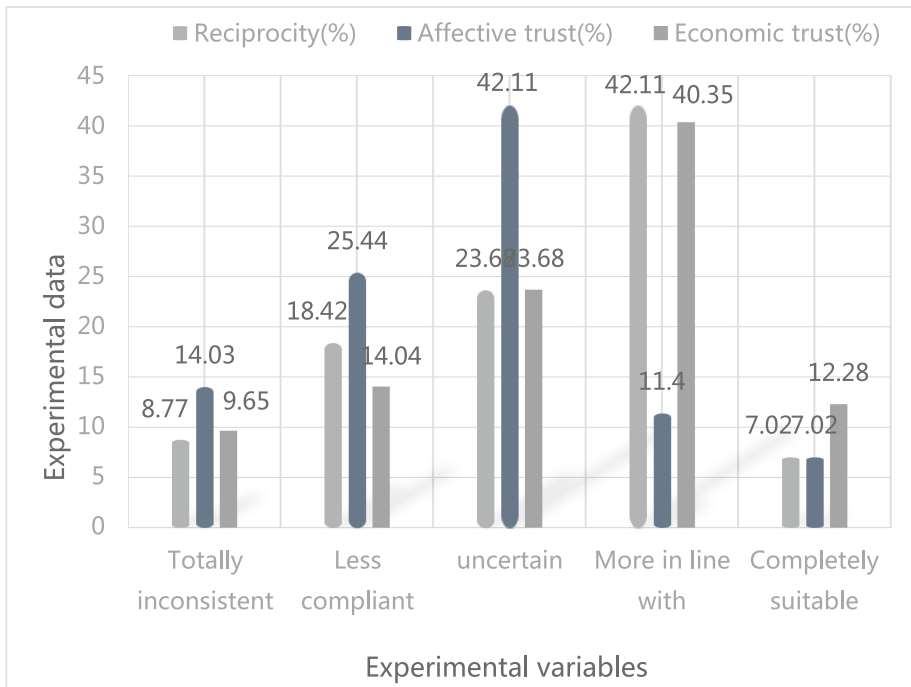


Fig. 2. Statistical analysis chart of study space atmosphere survey

As shown in Fig. 2, 49.13% of learners believe that when they help others solve problems, both the learner and the maker education resource sharing platform can achieve mutually beneficial help. Only 18.42% of learners believe that sharing knowledge in the maker education resource sharing platform can achieve the result of emotional trust. 52.63% of learners believe that the knowledge sharing behavior in the maker education resource sharing platform can achieve economic trust, which can save learning time and improve learning efficiency.

5 Conclusions

The research object of this paper is the design and research of the maker education resource sharing platform based on web technology. It can be obtained through experimental data. For learners, the learner can obtain what they want through the resource sharing platform. Of entrepreneurial knowledge and entrepreneurial experience and share their own feelings and experiences. Through an open and shared maker space, students' motivation for innovation and entrepreneurship can be greatly improved, and the advantages of shared resources can be maximized. This not only increases the importance of innovation to a greater extent, but also allows more Innovative entrepreneurs understand the industry thinking and characteristics of industry development. Secondly, it is possible to reconstruct the role and features of teachers to create a more professional and standardized team. Maker education is very different from traditional education. Therefore, learning content and learning methods must be changed. At the same time, teachers must also Innovate teaching techniques, strengthen innovation and entrepreneurship courses, and integrate them into innovative practice. In the cultivation of teaching goals, it combines educational reform and the cultivation of maker education talents to create creative talents with strong practical ability and rich theoretical knowledge.

References

1. Liu, H., Liu, Y.: Construction of a medical resource sharing mechanism based on blockchain technology: evidence from the medical resource imbalance of China. *Healthcare* **9**(1), 52–62 (2021)
2. Liu, S., Dai, Y., Cai, Z., et al.: Construction of double-precision wisdom teaching framework based on blockchain technology in cloud platform. *IEEE Access* **9**, 11823–11834 (2021)
3. Wang, P., Liu, L.: Strategy of library information resource construction based on FPGA and embedded system. *Microprocess. Microsyst.* **1**, 18–28 (2021)
4. Li, T.: Construction and implementation of network teaching platform for design art education based on cloud technology. *Educ. Sci. Theory Pract.* **18**(5), 149–182 (2018)
5. Hu, Y.: Construction of mental health education system in higher vocational colleges based on web platform. In: *E3S Web of Conferences*, vol. 218, no. 2, pp. 4–8 2020
6. Xiao, T.: Research on co-construction and sharing of higher vocational education information resources based on cloud computing. *Revista De La Facultad De Ingenieria* **32**(11), 984–989 (2017)
7. Luo, C., Chen, Y.: Design and research of private book sharing system based on sharing economy model. *J. Phys. Conf. Ser.* **1648**(4), 107–110 (2020)

8. Li, J.: Construction of modern educational technology MOOC platform based on courseware resource storage system. *Int. J. Emerg. Technol. Learn. (iJET)* **12**(9), 105–116 (2017)
9. Jing, L.: Construction of modern educational technology MOOC platform based on courseware resource storage system. *Int. J. Emerg. Technol. Learn.* **12**(9), 105–115 (2017)
10. Yan, L., Wang, H., Li, M.: Construction of an educational informatization platform based on P2P. *Agro Food Ind. Hi Tech* **28**(1), 371–375 (2017)
11. Zhou, Q., Lu, J., Yin, H., et al.: Layout guide for journal of physics: design and research of distributed educational resource sharing platform based on IPFS. *J. Phys: Conf. Ser.* **1971**(1), 55–56 (2021)
12. Yang, M., Jiang, P.: Improved Bayesian causal map approach for community-based product design project feasibility analysis. *IEEE Trans. Eng. Manag.* **1**, 1–19 (2019)



Travel Path Planning Algorithm Based on Deep Learning Technology

Chenglong Yu (✉)

School of Tourism and Humanities, Heilongjiang Polytechnic,
Harbin 150088, Heilongjiang, China

Abstract. Tourism resources are limited. In the process of development and utilization, they often cause a great waste. Deep learning technology can effectively integrate tourists with local natural landscape and cultural environment. This method realizes the processing of objective function value output results and calculation parameters by constructing a multi-dimensional data set with real-time and predictive characteristics, which contains rich semantic information, can be combined with other contents, and meet the use requirements of different demand objects at the same time. This paper mainly studies the visual dynamic interactive multi-attribute intelligent extraction algorithm based on deep learning technology. Firstly, the application status, existing problems and development status of deep neural network in the history at home and abroad are introduced. Secondly, the planning principle and characteristics of the algorithm, experimental results and theoretical model are analyzed. The experimental results show that when the parameter is 4, the error rate of the algorithm is the lowest, and the difference between the theoretical optimal path distance and the actual average optimal path distance is very small. Therefore, the algorithm can meet the needs of tourism path planning.

Keywords: Deep learning technology · Tourism path · Path planning · Tourism planning

1 Introduction

As an emerging industry, tourism has become an indispensable part of people's life, and many problems have arisen under the background of the rapid development of tourism [1, 2]. Tourism traffic is an important carrier for the development of tourism. Under the current situation of complex environment, huge data and huge amount of information, the traditional prediction based on artificial neural network can not meet the real-time processing of sea state changes and the influence of multiple factors. Therefore, deep learning technology came into being. As a new intelligent tool and method, deep learning technology (MIP) is widely used at home and abroad [3, 4].

Many scholars have studied path planning. Path planning is one of the key technologies of UAV autonomous flight. Traditional path planning algorithms have some limitations and shortcomings in complex dynamic environment. Some scholars have

proposed a deep reinforcement learning method for three-dimensional path planning using local information and relative distance without global information. In the actual scene with limited sensor capacity, UAV can obtain limited environmental information nearby. Therefore, path planning can be expressed as a partially observable Markov decision process [5, 6]. Autonomous Navigation in complex environment is an important requirement for robot planning. In order to improve the ability of autonomous navigation in complex environment, some scholars proposed a combination method of active slam (simultaneous and mapping) combining path planning and slam. He uses the full convolution residual network to identify obstacles and obtain the depth image. In the robot navigation, the double dqn algorithm is used to plan the obstacle avoidance path, and the two-dimensional map of the environment is constructed based on Fast SLAM. Experiments show that the algorithm can successfully identify and avoid different numbers of dynamic and static obstacles in the environment, and realize the autonomous navigation of robot in complex environment [7, 8]. The above research has laid an academic foundation for the research of tourism path planning algorithm based on deep learning technology.

This paper mainly studies the functional modules of automatic search, location and intelligent travel of scenic spots based on deep neural network structure and training set processing method of data. At the same time, aiming at the shortcomings of the traditional manual annotation algorithm, a new computing deployment method is proposed. Finally, the experimental results show that the system is of great significance to improve the development of China's tourism industry.

2 Discussion on Tourism Path Planning Algorithm Based on Deep Learning Technology

2.1 Tourism Route Planning

With the rapid growth of China's economy, tourism has become one of the indispensable, influential and competitive industries in the national economy. Under the background of the growing prosperity and development of the tourism market, the increasing number of tourists and the gradual improvement of people's living standards, there are many large-scale tourist attractions in more and more different regions or countries, or "scenic spot" scenic spots with concentrated tourist attractions but single business model. Most of these scenic spots take sightseeing as the main purpose and means to attract tourists for sightseeing, and have not formed a complete and systematic planning system and management methods. Tourism route planning is a way to manage the tourism market, but among all tourism activities, the most economic benefit is its regional characteristics. At the same time, one of the most important indicators to measure the basic operating capacity of a region's tourism industry is to evaluate whether it has the ability to plan tourism routes. Tourism route planning is the core of the operation and management of travel agencies. It is not only an integral part of tourism planning. There are few achievements in tourism route planning in this region, at home and abroad, and there are few high-level tourism route planning. With the development of tourism system, some basic features and characteristics of tourism leisure activities may be better understood

by people, and there may be more reasonable standards for tourism research in the future [9, 10].

The characteristics of tourism route planning are: (1) the variability of route planning: tourism route is a matter of great concern to tourists. Tourists usually have a general itinerary before going out, and tourists' demand for tourism services is becoming more and more popular. However, in reality, the travel plan may be affected by sudden changes in weather and traffic congestion, and the scenic spots and playing time need to be adjusted according to the actual situation. Therefore, there is some variability in the arrangement of tourist routes [11, 12].

(2) Complexity of resources: the tourism information elements involved in planning tourism routes are complex, including not only the individual factors of viewpoint distance, but also traffic information, tourism route information, landscape opening time, etc.

(3) The objective function is difficult to determine: the result of tourism route planning is difficult to be described by simple quantitative numerical indicators. Self cooking tourists should consider factors such as short travel time, low travel cost, short travel route, rich types of scenic spots included in the trip, etc. The planning of tourist routes must best meet the needs of tourists.

2.2 Deep Learning Technology

Deep learning is a sub field of machine learning. It learns multi-level representation and abstraction from data, and can solve supervised and unsupervised learning tasks. In short, it is a technology that can improve computer systems based on experience and data. Neural networks and many related skills create deep learning. Deep learning can be considered as a function with strong adaptability: by adjusting the parameters of the function, the output is infinitely close to the actual output of the sample. The deep learning model consists of several layers. The input of each layer is used as the output of another layer to realize the hierarchical representation of data.

Neural network is the basis of deep learning, also known as artificial neural network. Inspired by bionic research, it simulates the function of neurons in the brain. Neurons in the brain are composed of three parts: cell body, dendrites and axons. The function of cell body is to process signals. Dendrites are responsible for transmitting signals and axons are responsible for receiving signals. Hundreds of billions of these neurons form our brain through complex connections. It can process information, and its basic idea is to adapt to the connection between internal nodes according to the different complexity of the system. Although the great success of deep learning lies not only in imitating the function of the human brain, the mode of information transmission in the learning process of the human brain affects the initial development of the brain. The most primitive neural network model - MCP neuron model, its neural output is a binary variable:

$$y = \begin{cases} 0, & \text{other} \\ 1, & \sum_i w_i x_i \geq \theta \text{ and } z_j = 0, \forall j \end{cases} \quad (1)$$

Where, x_i is the input signal, w_i is the corresponding weight, and z_i is the inhibitory input. From the above formula, it can be seen that z_j can completely inhibit the excitation of neurons at any time, θ is the threshold neuron for excitation.

The goal of learning neural network is to find the parameters of the model and minimize the cost function. Gradient descent algorithm can solve this minimization problem well. Back propagation is an important method to calculate these gradients in neural networks. The basic idea of back propagation algorithm can be summarized as follows: firstly, the input layer data propagates layer by layer through the network until it passes through the output layer, and then the actual result and expected output result are obtained from the output layer, and the loss error between the actual result and expected result is calculated. Then, the calculated loss error is transferred layer by layer from the output layer to the input layer. In this process, the loss error is assigned to each layer of neurons, so that each neuron receives its own error signal, then calculates the gradient of the error signal, and uses these gradients to correct the weight of neurons. This process continues with the formation of the network until the network converges.

In order to calculate the gradient of the error signal, it is assumed that there is a network whose output value I of the j -th neuron in the L -th layer. Then it can be expressed as follows:

$$a = \partial \left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l \right) \quad (2)$$

Where w_{jk}^l represents the weight on the connection from the k -th neuron in the $(L-1)$ th layer to the j -th neuron in the L -th layer; Represents the offset of the j -th neuron in the L -th layer.

3 Experiment

3.1 SQLite Database

3.1.1 SQLite Structure

SQLite is a very popular lightweight database. It is stored in a relatively small C library that supports SQL, uses very few resources, and can run well with very little memory. Moreover, it is open source, supports multi language programming interfaces, and can be combined with c#, Java, PHP and other programming languages. Its main engine is independent. At present, it supports most operating systems. Anyone can use it without installing it. The structure of SQLite consists of four parts: SQL compiler, kernel, back end and accessories.

3.1.2 Geographic Location Data

Google maps is a map search application service launched by Google in China. Google Maps API is a set of application development interfaces provided to developers free of charge. API can provide tools and services such as basemap and panorama display, POI information retrieval, destination and route navigation, route planning, distance calculation and contact information conversion. Google maps has realized a large number of interactive map service functions in client websites, mobile device applications and servers. When accessing these interface services in their own applications, developers must first register and log in to their Google account, and then apply for and configure the developer key on the official website of Google Maps API.

3.2 Tourism Path Planning Process Based on Deep Learning Algorithm

Figure 1 is the flow chart of tourism path planning.

- (1) Parameters required for initialization algorithm: number of scenic spots, departure and arrival points, departure time, opening time of scenic spots, optimal stay time of scenic spots, scenic spot cost, travel time and travel budget.
- (2) Initialize the pheromone, calculate the heuristic path information, initialize the ant to the starting point and set the number of iterations g to 1.
- (3) Two ants select the nearest possible node j .
- (4) Judge whether the time $t_a(j)$ to reach the selected panoramic point J is appropriate. If $t_a(j) > TP(j)$, it means that the arrival time of tourists in the scenic spot is later than the closing time $TP(j)$, return to (3) and make another choice; If $t_a(j) < TP(j)$, the waiting time of tourist attraction J is calculated; If $t_a(j)$ is within the opening hours of the observation deck, there is no need to wait and you can visit it directly. Then continue to select the next attraction.
- (5) Calculate the travel time TT and cost TC , compare the routes RT and RC constructed by the time ant and cost ant, update the local pheromone according to formula (2), and then update the current optimal route of the global pheromone.
- (6) When all ants have built roads, calculate the pheromone increment of the overall optimal path, and update the overall pheromone according to formula (2).
- (7) Judge whether the current iteration number G is greater than g . if it is greater than g , stop the cycle and obtain the optimal result. If not greater than g , $g = G + 1$, turn to (3) to continue the iterative process.

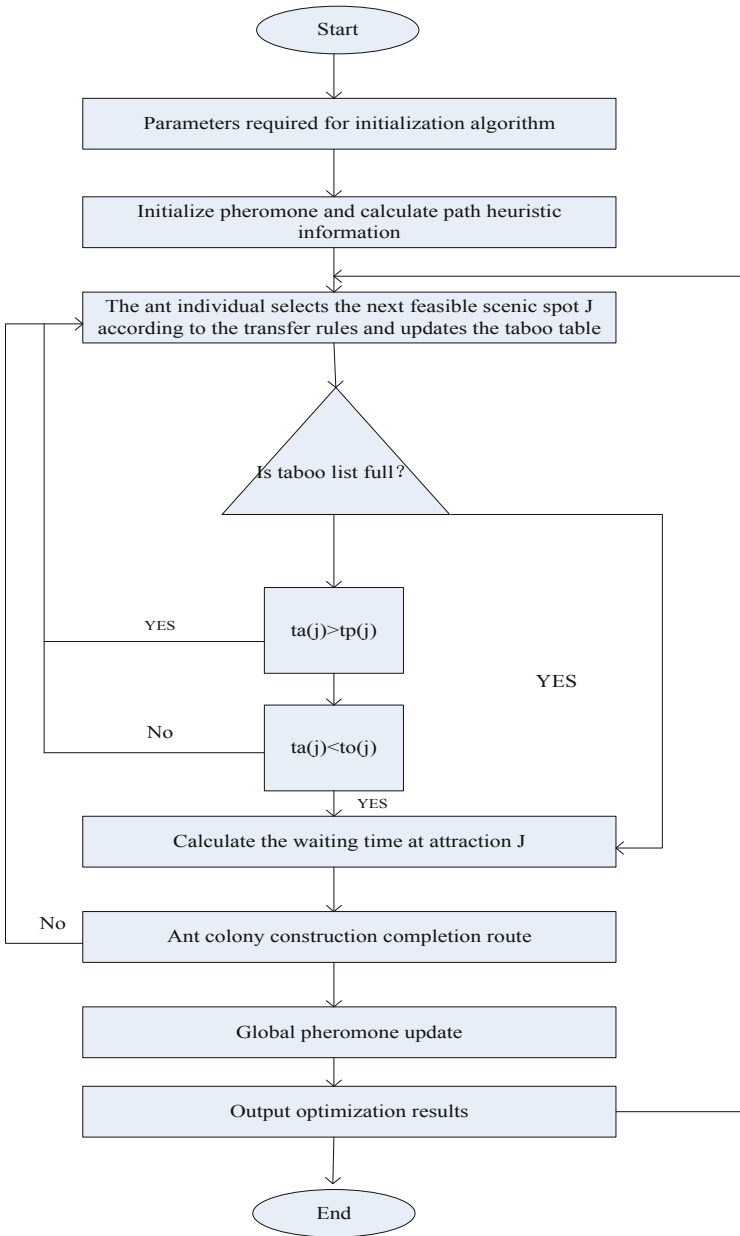


Fig. 1. Travel path planning process

4 Discussion

4.1 Optimal Path Distance Analysis

Table 1 lists the effects of different parameter values on the operation results.

Table 1. The effect of the parameters on the operation results

Parameter	Theoretical optimal path radial distance (km)	Average optimal path distance (km)	Error rate (%)
0	134376.9	132927.2	0.047%
1	132984.4	139874.1	0.063%
4	130943.7	129397.1	0.012%
2	130245.5	135298.7	0.029%
3	134457.1	137734.8	0.077%

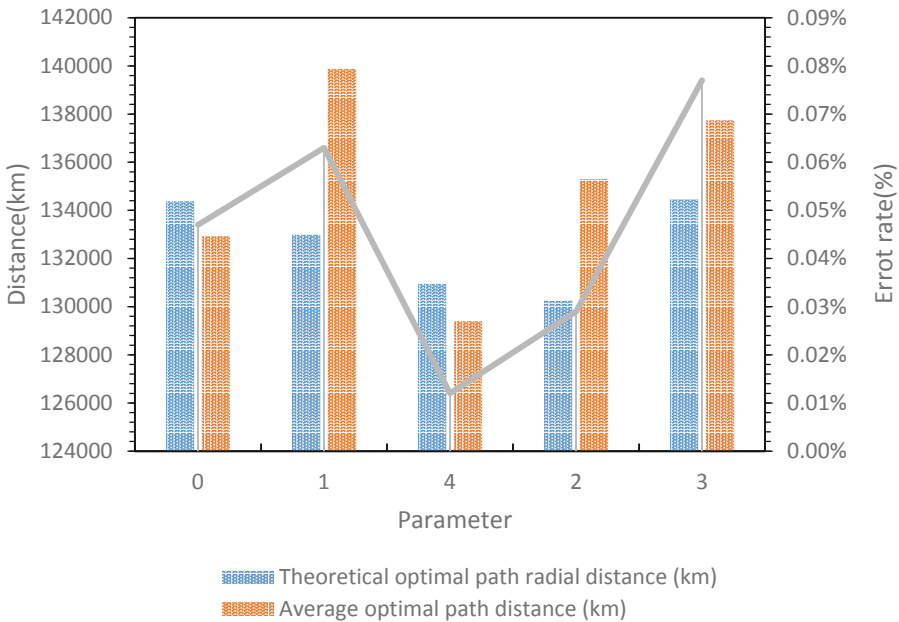


Fig. 2. Optimal path distance contrast map

This paper analyzes the operation results of the algorithm framework from three aspects: theoretical optimal path distance, average optimal path distance and error rate. As can be seen from Fig. 2, when the parameter is 4, the error rate of the algorithm is the lowest, and the difference between the theoretical optimal path distance and the actual average optimal path distance is very small. Therefore, the algorithm can meet the needs of tourism path planning,

5 Conclusion

Tourism transportation is the basis of tourism development. Therefore, this paper proposes a method to optimize the tour route of scenic spots in China based on deep learning algorithm. Firstly, the related technologies such as deep learning and artificial neural network are introduced. Secondly, the most important factors affecting the performance of path selection are analyzed. Then, combined with the current research status, future research direction and prospect, this paper analyzes the practical significance of the combination of deep learning and other automatic processing methods for improving path planning, and can solve the existing problems. Finally, it summarizes the content of the full text and points out the shortcomings, hoping that this paper can have further breakthrough and innovation in the field of tourism and transportation.

References

1. Xie, R., Meng, Z., Wang, L., et al.: Unmanned aerial vehicle path planning algorithm based on deep reinforcement learning in large-scale and dynamic environments. *IEEE Access* **9**, 24884–24900 (2021)
2. Wang, P., Li, X., Song, C., et al.: Research on dynamic path planning of wheeled robot based on deep reinforcement learning on the slope ground. *J. Robot.* **2020**(1), 1–10 (2020)
3. Qiu, L., Zhang, D., Tian, Y., Al-Nabhan, N.: Deep learning-based algorithm for vehicle detection in intelligent transportation systems. *J. Supercomput.* **77**(10), 11083–11098 (2021). <https://doi.org/10.1007/s11227-021-03712-9>
4. Yin, P., Zhang, L.: Image recommendation algorithm based on deep learning. *IEEE Access* **8**, 132799–132807 (2020)
5. Xu, S., Liu, et al.: Measuring benefits of Dianchi Lake wetland based on GIS RS GPS technology. *Ecol. Econ.* (2015-3), 230–235 (2021)
6. Kyungmee, P., Sangeun, L., Wang, S., et al.: 0426 sleep prediction algorithm based on deep learning technology. *Sleep (Suppl. 1)*, A172–A172 (2019)
7. Cheng, X., Qi, Y.: Indoor indicator path planning algorithm based on grid method. *Zhongguo Guanxing Jishu Xuebao/J. Chin. Inertial Technol.* **26**(2), 236–240 (2018)
8. Zhang, F., Li, N., Yuan, R., et al.: Robot path planning algorithm based on reinforcement learning. *J. Huazhong Univ. Sci. Technol. (Nat. Sci. Ed.)* **46**(12), 65–70 (2018)
9. Zhang, X., Lai, J., Xu, D., et al.: 2D lidar-based SLAM and path planning for indoor rescue using mobile robots. *J. Adv. Transp.* **2020**(3), 1–14 (2020)
10. Li, J., Chen, Y., Zhao, X., Huang, J.: An improved DQN path planning algorithm. *J. Supercomput.* **78**, 1–24 (2021). <https://doi.org/10.1007/s11227-021-03878-2>
11. Liu, Q., Shi, L., Sun, L., et al.: Path planning for UAV-mounted mobile edge computing with deep reinforcement learning. *IEEE Trans. Veh. Technol.* **69**(5), 5723–5728 (2020)
12. Zhou, X., Wu, P., Zhang, H., et al.: Learn to navigate: cooperative path planning for unmanned surface vehicles using deep reinforcement learning. *IEEE Access* **7**, 165262–165278 (2019)



Risk Forecast of International Engineering Project Based on Data Mining Technology

Depeng Li¹(✉), Wei Liu¹, and Rasha Almajed²

¹ China Construction Industrial & Energy Engineering Group Co., Ltd.,
Nanjing 210023, Jiangsu, China

² American University in the Emirates, Dubai, UAE

Abstract. Engineering construction is the foundation of the development of human society, and it is also the most important and dynamic field in the process of national economy and urbanization. During the period of rapid economic growth, various countries have put forward higher requirements for infrastructure and public facilities. An international engineering project is a complex system, which will be affected by various factors during the construction process. Therefore, it is very necessary to predict and manage project risks. In order to meet the needs of project operation and management, many large multinational groups in the world have established consulting companies that are in line with the world and are competitive, competitive and low-cost projects to predict project risks. At the same time, data mining technology has been well applied and developed in many fields, including risk assessment and risk prediction. Therefore, the progress and development of data mining technology provides a good technical support for the research on risk prediction of international engineering projects. This article adopts experimental analysis method and data analysis method, which is intended to combine data mining technology and related knowledge to establish a risk prediction model for international engineering projects, thereby reducing risks more efficiently. According to the experimental results, the expected output of the experimental model for the sample data is very similar to the actual output results, the gap is small, and the accuracy is relatively high.

Keywords: Data mining technology · International engineering projects · Risk factors · Risk prediction

1 Introduction

Project risk prediction is a dynamic process, so it is necessary to grasp the latest information at home and abroad in time. In actual construction, through the analysis of international engineering projects, we can understand the changing trends of the international market and domestic environment. Data mining is a processing technology that uses information. It can combine a lot of value without relying on existing knowledge and experience, and it can explain people's unknown problems or make predictive analysis of known things in the real world. At present, data analysis tools have gradually begun

to solve some practical problems in the fields of marketing and decision support. Therefore, combined with data mining, the research on the risk prediction of international engineering projects has important reference value.

At present, the research results of data mining technology in risk assessment and risk prediction are relatively rich. For example, Wang Huan pointed out that in-depth exploration of the principles and methods of applying database and data mining technology will help project risk management and provide reference value for further research on the application of data mining in project risk management [1]. Xiao Yibing believes that the distribution network fault prediction is an important work link in power grid engineering. In order to ensure the accuracy and scientificity of the distribution network fault prediction, it is necessary to pay attention to the application of data mining technology [2]. Zhang Chenyu introduced the necessity, process and main strategies of risk management for international engineering projects, and provided help and reference for risk management [3]. However, the research on the combination of data mining technology and international engineering project risk prediction is still relatively small. Therefore, this article will conduct research from this perspective, which to a certain extent has important practical significance and research value.

This article mainly discusses these aspects. First, the data mining technology and related theoretical basic knowledge are explained. Then, the risk prediction and related content of international engineering projects are introduced in detail. In addition, the application of RBF neural network in international engineering project risk prediction is also introduced. Finally, experimental research was carried out, based on data mining technology and related algorithms, an international engineering project risk prediction model was established, and test results and analysis conclusions were drawn.

2 Data Mining Technology and Risk Prediction of International Engineering Projects

2.1 Data Mining Technology

Data mining technology is a new, efficient and powerful information processing technology. It extracts a large amount of valuable information from a database or a group of storage media, and builds models based on these appearances to obtain potential user needs. It uses a large amount of information and knowledge to discover the characteristics of things in unknown fields, the laws of phenomena, and people make predictions about future development trends. At present, data mining technology is beginning to be used in many industries. For example, machine learning methods are applied to industrial production process automation and agricultural production management automation, etc. These tasks need to use this technology to analyze and sort out useful information.

Data mining, also known as knowledge discovery, is the process of discovering rules from a database. At present, this technology is a relatively young technology, and data mining has gradually become the subject of many internationally renowned scientific conferences and journals.

The development of data mining technology is based on traditional statistical theories and methods, including statistical analysis, neural network models, etc. The technology

of data mining mainly includes cluster analysis, neural network and association rule algorithm. Among them, cluster analysis divides different information data into multiple categories according to different classifications, such as univariate, multidimensional, and so on. Data mining is realized by clustering technology, which is divided into multiple small categories to classify and extract effective information. Association rule analysis methods and pattern recognition research are currently one of the most widely used and effective tools. When it is modeling things, it needs to build a large number of complex models to describe the interrelationships and internal regular relationships between things, which has some difficulties to a certain extent. Among all these data mining methods, association rule mining is the most active and widely used [4, 5].

2.2 Risk Prediction of International Engineering Projects

International engineering projects refer to the construction of engineering projects through international bidding, which are managed and constructed in accordance with international standardized management models, and the projects are undertaken by foreign contractors and the participation of companies from different countries. Due to differences in regional culture, political economy, and project construction funding sources, different international engineering projects are also very different.

The construction of international engineering projects is the most important way of the “going out” strategy at this stage. However, there are still many problems in the safety risk management of international engineering projects at this stage. These problems are concentrated in that they contain unsafe factors, which will affect the implementation and progress of international engineering projects at all times, and will affect the international engineering project itself and its consequences. The economic and social environment of the country or region where it is located has caused relatively large losses. The overall scale of international engineering projects is constantly increasing. At the same time, various uncontrollable risk factors are also increasing [6, 7].

Risk prediction is a very complex subject. Compared with national projects, due to differences in regions, languages, cultures, economics, and legal systems, the realization of project objectives for international projects is affected by many risk factors. The formation mechanism of project risk is shown in Fig. 1.

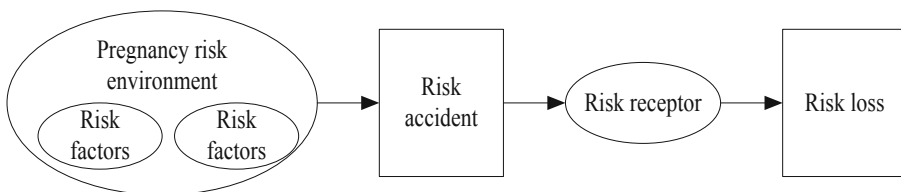


Fig. 1. Formation mechanism of international engineering project risk

The risks of international engineering projects are usually composed of risk accidents, factors and losses. Risk accidents are mainly events that deviate from the target due to the influence of a variety of risk factors. The occurrence of risk accidents is the main reason for risk losses. In simple terms, it means that the emergence of risks can only

be caused by risk accidents. This is also the main reason for risk losses. Risk accidents are often caused by a variety of different factors alone or together, and these factors themselves have certain contingency and regularity. By analyzing the law of occurrence of risk accidents, you can master the correct risk control strategy. It is also the focus of realizing risk early warning management and management.

Risk factors mainly refer to various factors that can lead to risk factors or enhance risk accidents. They are the most important cause of risk accidents and the most important cause of losses. There are also certain differences in risk factors for different risks, which can be distinguished as moral, material, and psychological risk factors. What needs to be paid attention to is that risk factors are the main cause of risk accidents and cannot be completely eliminated. Only through the implementation of corresponding measures can effective control of risk factors be achieved. Risk losses can be divided into indirect and direct losses. They mainly refer to the unexpected and unintentional economic loss caused during the period of risk [8, 9].

Compared with the process of general production activities, international engineering projects are affected by different risk factors, and their risks also have these characteristics. One is objectivity. Risk is not suppressed by human will. Their existence depends on the existence of risk factors. It happens with the emergence of risk factors. This is a common phenomenon in social practice. Therefore, people need to have an in-depth understanding of risks, record the rules of risk occurrence, and take appropriate risk management measures in a timely manner to remedy or reduce risks as much as possible. The second is relativity. The risk of a project is relative to the project object, and the same risk event has different impacts on different project objects. For example, the change of the engineering project is a risk for the customer, but it is an opportunity for the contractor, which can create conditions for the contractor to claim damages. On the other hand, the size of the project risk is also relative, the same risk has different degrees of tolerance to different problems. The risk tolerance of the project depends on factors such as investment scale, income level, project available resources and subject status. The third is integrity. The impact of risk is holistic and not limited to any specific period, place or aspect. For example, bad weather conditions will cause the project to stagnate, which will affect the later design of the entire project, which in turn affects the work of all participants in the later period. This will not only increase the cost, but also increase the construction time and seriously affect the quality of the project. Relatively independent local risks will accumulate and expand their impact as the project develops and time passes. The fourth is variability. Various risks will change in quantity and quality as the project progresses. Some risks disappear due to quick and efficient response methods, or are controlled over time, but there may also be new risks [10, 11].

2.3 Application of RBF Neural Network in Risk Prediction of International Engineering Projects

As one of the data mining techniques, RBF neural network can predict the potential losses that may occur or have occurred in the future by analyzing historical data and past experience, and it can also improve the predictability and reliability of the project by establishing a predictive model. It can be said that the RBF neural network has played an important role in constructing the risk prediction model of international engineering

projects, providing it with theoretical basis and technical support. The RBF neural network consists of three layers, and each layer has its own attributes that are different from other layers. The input layer is mainly composed of some sensory neurons. Its main function is to connect input variables and internal neurons so that they can transmit various information. It is only suitable for network buffering and connection [12]. Various input variables can be transmitted to neurons located in the hidden layer in turn. The specific calculation method is shown in formula (1) (2).

$$E(z) = \sum_{u=1}^B v_u \phi(\|z - d_u\|) \quad (1)$$

$$\phi(z, d_u) = f\left(\frac{-\|z - d_u\|^2}{\rho_u^2}\right), u = 1, 2, \dots, m \quad (2)$$

Among them, E is the network output layer selection function, x is the sample data, $\|\cdot\|$ is the norm, d is the center of the basis function, and v_u is the weight vector. ϕ is the radial basis function, B is the number of radial basis functions, and $\phi(\|z - d_u\|)$ is the set of B radial basis functions. ρ_u is the width of the hidden layer neuron basis function.

3 Experiment and Research

3.1 Experimental Environment

In this experiment, the tested model uses the C# language, calls Revit R API.dll in the Visual Studio 2019 environment, and runs the external runtime program (warning module) from an external (external control) command. Due to limited programmability, the model can be used to independently call and process data in the Revit SDK, and it can also realize the early warning function. And use PAI interface, Revit SDK data package, Visual Basic.net, C#, C++, etc. Using the Revit DB Link plug-in and linking the plug-in to the Revit main database, the original data of the risk prediction model of international engineering projects can be directly imported into external databases such as Microsoft Access, Microsoft Excel, and ODBC databases through the plug-in.

3.2 Experimental Process

In this experiment, based on data mining data mining technology and related algorithms, an international engineering project risk prediction model is constructed. After inputting the sample data Sample 1, Sample 2, Sample 3, and Sample 4 in the model, the expected output and actual output results are obtained by testing respectively. And analyze the expected output and actual output results, and draw conclusions. Some experimental results are shown below.

4 Analysis and Discussion

In this experiment, data mining technology and related algorithms are introduced to establish a predictive model. After the sample data is input, the expected output and actual output are obtained by testing. The test results are shown in Table 1.

Table 1. Test results of the model

Project	Sample data input	Expected output	Actual output
Sample 1	0.29	0.31	0.323
Sample 2	0.35	0.39	0.385
Sample 3	0.45	0.51	0.503
Sample 4	0.37	0.41	0.397

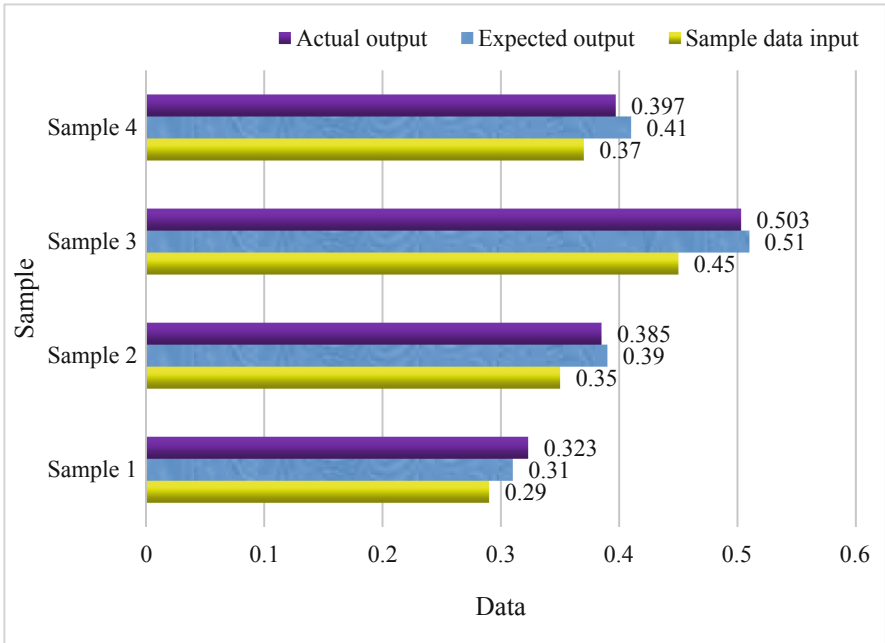


Fig. 2. Test results of the model

It can be seen from Fig. 2 that the prediction model established by the introduction of data mining technology and related algorithms, after inputting sample data, tests to obtain the expected output and actual output. The expected output of Sample 1, Sample 2, Sample 3, and Sample 4 are 0.31, 0.39, 0.51, and 0.41, respectively, and the actual output is 0.323, 0.385, 0.503, and 0.397. It can be seen that the expected output of the experimental model for the sample data is very similar to the actual output result, the gap is small, and the accuracy is relatively high.

5 Conclusion

Data mining technology has a wide range of applications, including engineering design, management and decision-making. At the same time, in the process of international

engineering projects, it is necessary to conduct risk assessment and risk prediction, but the large amount of data and information involved is very large. Therefore, in the process of project management, it is very necessary to use data analysis, processing and other means to predict risks. In this process, it is necessary to take advantage of data mining technology to collect information and materials required at various stages for screening and sorting, and then after analysis, evaluation and judgment, it needs to be converted into usable resources and finally realized risks prediction. Therefore, this article focuses on the combination of data mining technology and international engineering project risk prediction.

References

1. Wang, H.: Research on project risk management based on data mining technology. *Inf. Rec. Mater.* **18**(005), 94–96 (2017)
2. Xiao, Y.: Research on distribution network fault prediction based on data mining technology. *Equip. Manag. Maintenance* (2018-17), 35–36 (2021)
3. Zhang, C.: Analysis and research on risk management of international engineering projects. *Foreign Econ. Trade* (2013-1), 100–103 (2021)
4. Ozkan, M.B., Karagoz, P.: Data mining-based upscaling approach for regional wind power forecasting: regional statistical hybrid wind power forecast technique (RegionalSHWIP). *IEEE Access* **7**, 171790–171800 (2019)
5. Liang, H., et al.: Data mining-based model and risk prediction of colorectal cancer by using secondary health data: a systematic review. *Chin. J. Cancer Res.* **32**(02), 124–133 (2020)
6. Kwilinski, A., Litvin, V., Kamchatova, E., et al.: Information support of the entrepreneurship model complex with the application of cloud technologies. *Int. J. Entrep.* **25**(1), 1–8 (2021)
7. Delpla, I., Florea, M., Rodriguez, M.J.: Drinking water source monitoring using early warning systems based on data mining techniques. *Water Resour. Manag.* **33**(1), 129–140 (2018). <https://doi.org/10.1007/s11269-018-2092-4>
8. Howard, C., Hernandez, M.K., Livingood, R., et al.: Applications of decision tree analytics on semi-structured north atlantic tropical cyclone forecasts. *Int. J. Sociotechnology Knowl. Dev.* **11**(2), 31–53 (2019)
9. Jeon, J.H., Kim, D.N., Lee, S.H.: Analysis of experience knowledge from defense M&S based on data mining technique: the case of defense M&S for education and training. *J. Korea Technol. Innov. Soc.* **23**(6), 1202–1221 (2020)
10. Yoon, B., Jeong, Y., Kim, S.: Detecting a risk signal in stock investment through opinion mining and graph-based semi-supervised learning. *IEEE Access* **8**, 161943–161957 (2020)
11. Abubakari, M.S., Arifin, F., Hungilo, G.G.: Predicting students' academic performance in educational data mining based on deep learning using TensorFlow. *Int. J. Educ. Manag. Eng.* **10**(6), 27–33 (2020)
12. Rejaul, M., Royel, I., Jaman, M.A., et al.: Machine learning and data mining methods in early detection of stomach cancer risk. *J. Appl. Sci. Eng. Technol.* **24**(1), 1–8 (2020)



Speech Recognition Technology of English Translation System Based on DTW Algorithm

Peilan Peng^{1,2}(✉)

¹ ZHIYUAN (Liberal Study) College, Beijing Institute of Petrochemical Technology, Beijing 102617, China

² National Research Center for Foreign Language and Education, Beijing Foreign Studies University, Beijing 100089, China

Abstract. Translation is an ancient and highly practical subject. With the development of the times, people pay more and more attention to language. As one of the most important tools in international communication, English has become a topic of common concern for people all over the world. Under the modern language system, it has become very necessary to study multiple strategies to replace traditional sentence patterns or morphological words. At the same time, in the era of big data, translation technology is one of the important components. Among them, multi-strategy translation uses certain technical means to translate the original text into the target language in different language environments. It can enable the target reader to quickly understand and accept the intended information or intention, and can also quickly and accurately express the original content and obtain accurate results. The DTW algorithm is a shortest path method based on the minimum non-random search method. It has strong robustness when solving multi-target translation problems and can well overcome many shortcomings in traditional classical grammatical analysis. This article adopts experimental analysis and data analysis methods to explore the multi-strategy English translation system by combining the research of DTW algorithm, so as to achieve high-quality and high-efficiency translation of the target language. According to the experimental results, the recognition time of the system tested in this experiment is relatively short, and the recognition rate can be as high as 98.3%, and the work efficiency is relatively high.

Keywords: DTW algorithm · English translation system · Best path · Speech recognition

1 Introduction

With the rapid development of modern science and technology, social economy and cultural life, people use English more and more widely in daily communication. As an important language tool and information processing method, the multi-strategy translation system has also become a research topic in the field of modern society translation and other fields. The DTW algorithm is a multi-parameter calculation method based on

the least squares method. It converts the function optimization problem into a linear programming problem, and obtains the best performance by solving the optimal solution. The basic idea is to use the simplest global search capability to make the shortest path, maximize the goal and minimize the cost. Combining the research of DTW algorithm to explore the multi-strategy English translation system, so as to achieve high-quality and high-efficiency translation of the target language.

From the current point of view, the research results on DTW algorithm, English translation strategy and machine translation in academia are relatively rich. For example, Wen Yuhua proposed that compared with the traditional correction system, the automatic correction system for English pronunciation errors based on the DTW algorithm optimizes the speech recognition sensor and the pronunciation recognition processor, which makes the accuracy of the English pronunciation correction higher [1]. Hong F believes that in English, the use of prepositions is more flexible and common. According to the differences between English and Chinese languages, the translation strategies of prepositions in translation between English and Chinese are analyzed and discussed [2]. WANG Yu-han pointed out that the rapid development of information technology in the past ten years has made people more aware of the importance of translation. Combining the relevant theories of multimodal discourse analysis to explore the concept of multimodal teaching is conducive to mobilizing students' enthusiasm and improving the efficiency and quality of teaching [3]. Since there are few researches on the combination of DTW algorithm and multi-strategy English translation system at present, this article will combine this algorithm to carry out research in this area.

This article mainly discusses these aspects. First, the DTW algorithm and its related theoretical knowledge are explained. Then, it elaborates on the multi-strategy English translation system and its related aspects. In addition, the application of DTW algorithm in multi-strategy English translation system is also introduced. Finally, based on the DTW algorithm, an experimental study on the multi-strategy English translation system is carried out, and the experimental results and analysis conclusions are drawn.

2 DTW Algorithm and Multi-strategy English Translation System

2.1 DTW Algorithm

DTW algorithm, short for dynamic time warping algorithm, has achieved great success in the recognition of isolated words with small vocabulary. It adjusts the pronunciation speed and re-adjusts the characteristic parameter sequence mode to introduce as few errors as possible to obtain the best results. He has achieved great success in recognizing individual words with a small vocabulary. Due to changes in speech rate, intonation, and speaker's accent, the recognition module and the reference module cannot be completely consistent over time. This algorithm can compare the two at the equivalent point in time, and solve the problem of unequal speaker speed [4, 5].

Dynamic programming is to turn a multi-step process into a series of single-step problems, solve them one by one, and optimize the problems in the solution process. It can effectively find the best solution for many overlapping sub-problems. It divides the

problem into sub-problems and uses the principle of optimality to establish a recursive formula to calculate the optimal solution. In order to avoid the repeated solution of these sub-problems, their results are gradually calculated and recorded, from simple problems to overall problems to be solved. In other words, dynamic programming records the results of recursion so that it does not take time to solve the same problem.

The essence of the DTW algorithm is to use the idea of dynamic programming and use local optimization to automatically find a path. Along this path, the cumulative distortion between the two feature vectors is the smallest, that is, to find the best path through each intersection from the starting point to the ending point of a point, so that the total frame distortion of all intersections on the path is minimized. In the learning phase of the system, the user saves the feature vector sequence of each word in the vocabulary as a template in the template library. In the recognition step, the feature vector sequence of the input speech is sequentially compared with each model in the model library, and the one with the highest similarity is output as the recognition result. Dynamic time warping is the use of dynamic programming technology to transform complex global optimization problems into many local optimization problems and make incremental decisions [6, 7].

2.2 Multi-strategy English Translation System

Translation is an important means of language information exchange. In the process of daily communication, business negotiation or other communication, people use English to facilitate communication and express their needs.

Early machine translation systems are generally based on the analysis of part of speech and word order, and statistical techniques are often used in the analysis. People are increasingly finding that the communication between different languages is becoming more and more common in their study, life and work. Whether it is idioms or written communication, there are stricter, more urgent, and more demanding requirements for English translation. On the other hand, computing power is also advancing by leaps and bounds. The development and popularization of the Internet, the multilingual archives of bilingual countries, the United Nations, etc. have enabled people to use tens of millions of bilingual parallel corpus sentences, and the corpus is constantly expanding. All of these provide the necessary theoretical and practical basic knowledge for the study of multi-strategy English translation systems [8, 9].

The “Pharaoh” translation system belongs to the earlier machine translation system. It mainly consists of two parts, namely training and decoding. For the input sentence, the sentence is divided first, and then the best combination is found according to the existing sentence table to generate the final translation. Because its principle is simple and easy to use, many researchers have experimented with it or compared it with the basic system. Therefore, its appearance has played a very important role in promoting the development of statistical machine translation. The structure of the “Pharaoh” translation system is shown in Fig. 1.

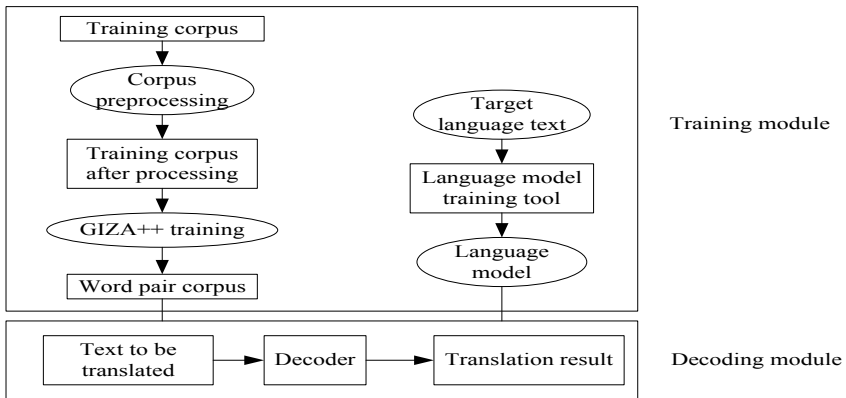


Fig. 1. "Pharaoh" translation system structure diagram

Machine translation methods are divided into rule-based translation methods and corpus-based translation methods. Machine translation refers to the use of computers to translate one natural language (source language) into another natural language (target language). Initially, due to the slow speed and low accuracy of machine translation, the cost of machine translation was much higher than that of manual translation. This has had a huge impact on the prospects of machine translation research, and machine translation research has declined. Later, with the rapid development of science and technology, cultural exchanges between different ethnic groups became more frequent and language barriers became more serious. People are beginning to realize that machine translation has a huge demand and application potential.

At present, statistical machine translation is a research site in the field of machine translation and even natural language processing. The multi-strategy translation system is a complex process, the core of which is to translate and integrate the original text at multiple levels, and translate the translation into the target language. Multi-strategy translation system is a relatively advanced technology and method. It is mainly divided into three parts in the translation process. The first step is to combine the words with the target corpus. At this stage, a large amount of data information is needed to analyze, organize and calculate the content to be decoded. The processing mechanism of the system refers to the comprehensive analysis of multiple different types of texts to achieve the optimal combination and convert the best translation into the most ideal expression effect [10, 11].

Multi-strategy translation should follow these two basic principles. One is faithful to the original text. Before extracting text information, it is necessary to have a correct understanding of the content of the original text and the author, the target reader and the identity of the translator contained in the translation, etc., so as to ensure the translation of high-quality articles. The second is accuracy. It is necessary to accurately express the target semantics and make it match the original intent. At the same time, it should be noted that there can be no major differences between different methods.

2.3 Application of DTW Algorithm in Multi-strategy English Translation System

The DTW algorithm is a nonlinear warping technique that combines time warping and distance measurement. In speech recognition, due to the considerable randomness of the speech signal, the input pattern cannot be simply compared with the corresponding reference pattern, but even if the same sound is made by the same person at different times, it may not be exactly the same. Using this algorithm to re-adjust the parameter attributes of the speech signal can solve the problem of two vector lengths with different lengths, so as to achieve better results in the speech recognition of the English multi-strategy translation system. For simplicity and processing, the Euclidean distance or its twisted form is usually used for measurement. Then find the cumulative distance of the best path, and finally get the corresponding frame matching distance to maintain a better path [12]. The calculation method is shown in formula (1) (2) (3).

$$f[R(b), Q(a)] = \sum_{b=1}^o (r_b - q_a)^2 \quad (1)$$

$$F[(b_u, a_u)] = f[R(b_u), Q(a_u)] + F[(b_{u-1}, a_{u-1})] \quad (2)$$

$$F[(b_{u-1}, a_{u-1})] = \min\{F[(b_{u-1}, a_u)], F[(b_{u-1}, a_u - 1)], F[(b_{u-1}, a_u - 2)]\} \quad (3)$$

Among them, a is the time sequence label of the training speech frame, b is the time sequence label of the test speech frame, R is the test mode, and Q is the reference mode. $F[R, Q]$ represents the distortion between the test pattern and the reference pattern, and o is the dimension of the feature vector. $T(b)$ and $Q(a)$ are both speech frames, $T(b) = \{r_1, r_2, \dots, r_p\}$ and $Q(a) = \{q_1, q_2, \dots, q_p\}$ respectively represent $T(b)$ and $Q(a)$ feature vectors.

3 Experiment and Research

3.1 Experimental Environment

In this experiment, the online translation services provided by Baidu and Google and the open source statistical machine translation system were used to translate the experimental samples from English to Chinese to obtain the corresponding translations. Among them, the operating system uses Windows 10, the development tool uses Python IDLE, and the development language is Python. The operating mode is C/S structure, namely: Client/Server (client/server) structure. Hardware configuration: CPU is Inter i5 2.9 GHz, memory is 16 GB. Software configuration: Web server is Tomcat6.0, background database is MySQL5.9, development and debugging environment is Eclipse4.2. Network configuration: bandwidth is 1000 Mbps for wired.

3.2 Experimental Process

In this experiment, the system is comprehensively debugged, which includes training each control sample separately, using the training result as a template for pattern matching, and finally getting the recognition result. The specific operation performance is as

follows: Regarding system performance issues, the samples 1, 2, 3, and 4 are tested in terms of the longest recognition time, the shortest recognition time, and the recognition rate. Some experimental results are shown below.

4 Analysis and Discussion

In this experiment, in terms of system performance issues, Sample 1, Sample 2, Sample 3, and Sample 4 were tested in terms of the longest recognition time, shortest recognition time, and recognition rate. The experimental results are shown in Table 1.

Table 1. System performance test results

Project	Sample 1	Sample 2	Sample 3	Sample 4
Number of experiments	50	50	50	50
Maximum recognition time (ms)	78	96	67	88
Shortest recognition time (ms)	57	73	46	51
Recognition rate (%)	97.3	96.8	98.3	97.9

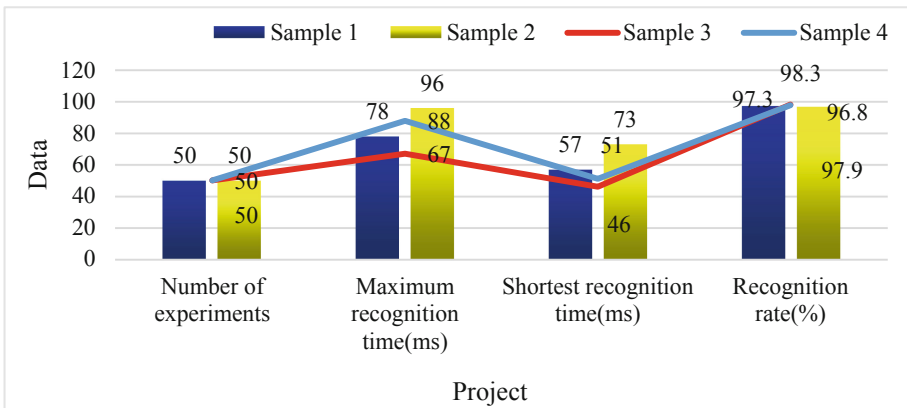


Fig. 2. System performance test results

It can be seen from Fig. 2 that the shortest recognition times of Sample 1, Sample 2, Sample 3, and Sample 4 are 57 ms, 73 ms, 46 ms, and 51 ms, respectively, and the recognition rates are 97.3%, 96.8%, 98.3%, and 97.9%, respectively. It can be seen that the recognition time of the system tested in this experiment is relatively short, and the recognition rate of up to 98.3% can be achieved, and the work efficiency is relatively high.

5 Conclusion

In recent years, as international exchanges have become more frequent and information exchanges have become increasingly globalized. This makes people gradually produce new requirements for the language environment. Therefore, how to accurately and efficiently translate a suitable translation while satisfying multiple functional requirements has become an important issue facing linguistics. With the progress of society, people have higher and higher requirements for computer systems, and multi-strategy translation systems are increasingly used in various fields, which makes the research of translation systems a hot spot. Therefore, this article is based on the DTW algorithm to study the multi-strategy English translation system, which has certain practical significance and research value for the future research on the English translation system.

References

1. Wen, Y.: Design of automatic correction system for English pronunciation errors based on DTW algorithm. *Mod. Electron. Technol.* **43**(561(10)), 132–134 (2020)
2. Hong, F., Chen, J., Zhang, Z., et al.: Time series risk prediction based on LSTM and a variant DTW algorithm: application of bed inventory overturn prevention in a pant-leg CFB boiler. *IEEE Access* **8**, 156634–156644 (2020)
3. Wang, Y., Yan, Z., Wu, T., et al.: Analysis of infant speech based on DTW algorithm. *J. Qiqihar Univ. (Nat. Sci. Ed.)* **035**(004), 69–73 (2019)
4. Lee, C.S., Hou, H.T., Jiang, W.W.: A DTW-based fast searching heuristic with application on establishing a frozen shoulder rehabilitation system. *Appl. Sci. Manag. Res.* **6**(1), 1–17 (2019)
5. Tong, X., Yang, M., Zhang, G.: A protection scheme for HVDC transmission line based on traveling waveform similarity using improved DTW algorithm. *Zhongguo Dianji Gongcheng Xuebao/Proc. Chin. Soc. Electr. Eng.* **40**(12), 3878–3887 (2020)
6. Wang, K.S., Joo, Y.H.: Tracking of multiple moving objects using DTW algorithm. *Trans. Korean Inst. Electr. Eng.* **68**(5), 670–677 (2019)
7. Hong, Z., et al.: A DTW distance-based seismic waveform clustering method for layers of varying thickness. *Appl. Geophys.* **17**(2), 171–181 (2020). <https://doi.org/10.1007/s11770-020-0819-5>
8. Yuan, Z., Jin, C., Chen, Z.: Research on language analysis of English translation system based on fuzzy algorithm. *J. Intell. Fuzzy Syst.* **40**(3), 1–9 (2020)
9. Chatterjee, N., Gupta, S.: Efficient phrase table pruning for Hindi to English machine translation through syntactic and marker-based filtering and hybrid similarity measurement. *Nat. Lang. Eng.* **25**(PT.1), 171–210 (2019)
10. Ban, H., Ning, J.: Design of English automatic translation system based on machine intelligent translation and secure internet of things. *Mob. Inf. Syst.* **2021**(7639), 1–8 (2021)
11. Wen, H.: Intelligent English translation mobile platform and recognition system based on support vector machine. *J. Intell. Fuzzy Syst.* **38**(153), 1–12 (2020)
12. Hannan, A., Sarma, S.K.: Marie: a statistical approach to build a machine translation system for English assamese language pair. *Int. J. Comput. Sci. Eng.* **7**(3), 774–779 (2019)



Design and Implementation of the Life Cycle Management System of Power Tools Based on the Internet of Things

Zhengfeng Xie¹(✉), Hongyun Shi¹, Lianggang Xu¹, Shichun Wang¹, Shi Qiu¹, Long Yang¹, and Rajan Singh Tiwari²

¹ Transmission Operation and Maintenance Branch of Guizhou Power Grid Co., Ltd., Guiyang 550002, Guizhou, China

² Technocrats Institute of Technology, Bhopal, India

Abstract. With the continuous development of society and economy, the scope of transmission lines continues to expand, the traffic of transmission line inspections is increasing day by day, and the frequency of use of safety tools is also increasing. At the same time, the traditional manual management of safety tools and equipment due to complicated steps, long cycles and low execution efficiency can no longer meet the needs of modern transmission line inspections for safety tools and equipment management. It is an urgent task to develop a whole life cycle management system for safety tools and equipment using new technology. This article aims to study the design and implementation of a full life cycle management system for power tools and appliances based on the Internet of Things. Based on the analysis of QR code technology, system performance requirements, system function design and QR code related algorithms, the basic management of the system. Safety tools and equipment life management, safety tools and equipment maintenance management, and safety tools and equipment query statistics modules were designed, using QR code technology to realize the automatic identification of safety tools and equipment, and then realize the purchase, warehousing, and outgoing of safety tools and equipment. Informatization management for the entire life cycle of inspection and scrapping, improve the management level of safety tools and equipment warehouses of transmission lines, eliminate irregular operations, and ensure the smooth development of safe production of transmission lines.

Keywords: Internet of Things · Power tools · Full life cycle management · System design

1 Introduction

The power industry is one of the industries that pay great attention to safety, and the management of safety tools and equipment is an important part of the daily safety production management of the power industry [1, 2]. For the safe and stable operation of transmission lines, operation and maintenance personnel will use various safety devices

to conduct inspections on power lines and tower equipment during daily inspections [3, 4]. Due to the wide variety and large number of equipment contained in safety tools and appliances, their management involves many aspects such as types, quantities, inspections and tests, scrapping and renewal, and there are greater difficulties in management. Therefore, the information management of the quality, service life, replacement, etc., of these safety tools has become very important.

The main purpose of power companies to manage safety tools is to provide better quality tools for production activities, help operation and maintenance personnel to carry out daily inspections of transmission lines in a safer and more reliable way, and to maintain production in a safe and stable manner [5, 6]. Under the current state, it promotes the continuous improvement of the safety of production activities and helps power companies achieve greater results in terms of economic and social benefits [7, 8]. Therefore, it is of great significance to strengthen the management of safety tools and equipment. The management level of power companies will directly affect their actual production level and actual management capabilities. Therefore, power companies strengthen the management and monitoring of safety tools and equipment, and use information technology to improve management quality and efficiency [9, 10]. Not only can they help power companies improve their modern management level, effectively promote safe production, but also save labor costs and use information technology instead of low effective manual management mode improves management efficiency [11, 12].

This article sorts out the requirements for the full life cycle management of safety tools and equipment. Based on the QR code technology and the Internet of Things technology, with the help of software engineering related theories and specific methods to research, establish a reliable and practical safety tool management. The system includes analyzing system requirements, designing specific functions, and carrying out project realization and specific testing.

2 Design

2.1 System Performance Requirements

(1) System reliability

From the perspective of non-functional requirements, the system cannot be verified in the early design and implementation of the system, and the stage of online operation. The concurrent system data and the number of users may not reach a certain level, so in the process of designing the system architecture. It is necessary to fully consider various factors in order to determine the design mode with a more reasonable structure.

(2) System operability

Ordinary users are not professional computer practitioners, so the operability of the software interface must be high, and an intuitive and clear operation interface must be presented to the user. At the same time, it must ensure a faster response time and ensure that multiple users can use it at the same time.

2.2 System Function Design

The specific functions of this system include four functional modules: system management, safety tools and equipment full life cycle management, safety tools and equipment maintenance management, and safety tools and equipment query statistics. The module diagram is shown in Fig. 1.

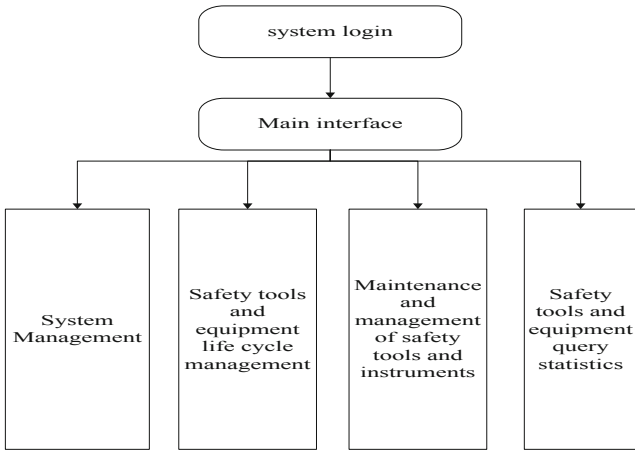


Fig. 1. System function architecture

2.2.1 System Management Module

System login refers to the control of user login status by verifying user name and password and checking its specific authority. Can be divided into account authority management, user management and login management. Authority management refers to the assignment of specific authority of users, adding, deleting and changing the specific authority of users. User management refers to adding, deleting and changing user names and passwords in the system.

2.2.2 Full Life Cycle Management Function Module

The whole life cycle of safety tools and equipment is the core function of this system, which has modules such as tool and equipment procurement, tool and equipment distribution, tool and equipment storage, tool and equipment testing and maintenance, and tool and equipment scrap management. Specific as shown in Fig. 2.

(1) Procurement of tools and equipment

The management of safety tools and instruments should start with the procurement plan. According to the configuration, use and inventory of the overall safety tools and equipment in the previous year, the user shall apply for the summary and formulate the annual procurement plan. After the procurement plan is reported

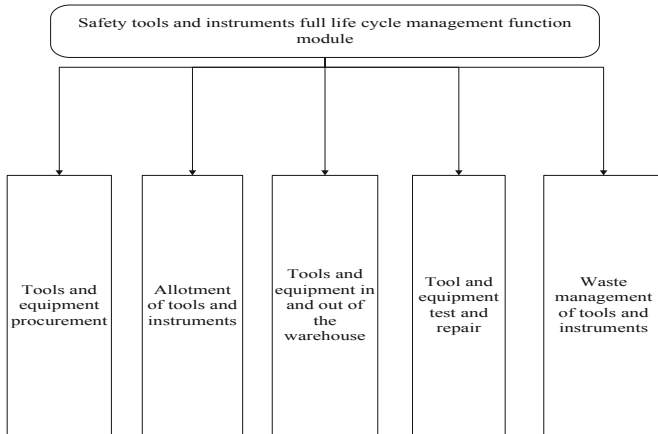


Fig. 2. Full life cycle functional architecture of safety tools

to the superior management unit for budget approval (may be adjusted), enter the procurement procedure according to the approved plan (independent procurement, bidding procurement or outsourcing procurement, etc., this process is not managed in this system). After the procurement is completed, all equipment enters the test center. In the test area, the first inspection of tools and appliances is prepared. For tools and appliances that pass the first inspection, they will be identified by a two-dimensional code, and then put into storage according to regulations.

(2) Allotment of tools and instruments

Safety tools and appliances are managed in a secondary warehouse, that is, all tools and appliances are purchased in the warehouse management center, and the power supply stations and other users will collect them from the central warehouse according to the application or plan. The use stage is stored in the warehouses of the power supply bureaus. The tools and instruments distributed to various departments are returned to the central warehouse through the process of testing, overhauling, and scrapping.

(3) In and out of warehouses

During the operation of lending and returning safety tools and instruments, the QR code identification records corresponding to the safety tools and instruments must be scanned as required, and unified and standardized information records are realized for the outbound and in-warehouse. Only the instruments and instruments that have been scanned and identified can be loaned and returned. At the time, the integrity check operation before the return must be followed, such as return photos, etc., before the return can be completed, forming a complete closed-loop management process of tools and instruments in and out of the warehouse.

(4) Test and maintenance of tools and instruments

According to the latest test date, test period and other information in the historical archives of safety tools, compile test and inspection plans for safety tools and appliances, and complete inspection and test management. After the test and inspection are completed, directly enter the central warehouse, and then the user

will receive it; according to requirements, the management unit will not be changed during the test, and the user will receive it after the test is completed.

(5) Waste management of tools and instruments

For tools and instruments that are unqualified for testing and maintenance or damaged in use, they need to be recorded in the year-end or regular scrapping module. The scrapped tools and instruments are removed from the inventory, and the subsequent processing is the responsibility of the asset management department and is not managed in this system.

2.2.3 Safety Equipment Maintenance Management Sub-module

The specific maintenance work of safety tools and appliances is automatically assigned to the warehouse safety tools and appliances manager to complete on time according to the maintenance, inspection, and test cycles of different tools and appliances. Daily maintenance work must be completed within the specified time, and the maintenance work must be completed on the safety tools and equipment maintained in accordance with the specific steps. The focus of the management work is to complete the maintenance work within the maintenance period, and not to omit the safety tools and equipment during maintenance. Only in this way can the safety tools and equipment themselves be guaranteed and meet the actual needs in the production work.

2.2.4 Query Statistics of Safety Tools and Instruments

Through the system for transparent management of various processes in the entire life cycle of tools and appliances, the status of tools and appliances can be based on evidence. In the whole life cycle management of tools and appliances, tools and appliances will involve purchasing, allotment, warehousing, testing and maintenance, and scrapping. Through the system's online control of the process of tools and equipment, the status of chemical equipment can be transparent. Managers will be able to intuitively grasp the status of tools and equipment from multiple dimensions, and automatically record the entire process with informatization, completely changing the management mode of power safety tools and equipment, and achieving lean management.

2.3 Internet of Things Technology

The two-dimensional code of the Internet of Things is a bar code that uses a certain geometric figure to record data information in a spaced pattern distributed on a plane according to a certain rule. Between the one-dimensional bar code RFID technology. It has a large amount of information, has the functions of error correction and encryption, and has a wide range of applications. The production of a two-dimensional code is different from traditional information carriers. It is a graphic data file that does not have strict requirements on the medium. It can be embodied in a variety of materials, which is convenient for information transmission and real-time detection of equipment. The two-dimensional code can realize the reliable storage of various large-capacity information, and is convenient to carry. It can hold up to 1,850 uppercase letters or 2,710 numbers or about 500 Chinese characters in bytes. It is a major technology in a barcode. The

two-dimensional code is added in the design of the replacement operation under the condition of lossless and misplaced. The encryption process is as follows:

- (1) First select two large prime numbers p and q . It is required that p cannot be equal to q , and there is a certain gap between p and q .
- (2) Calculation

$$n = p * q \quad (1)$$

- (3) Calculate

$$\Phi(n) = (p - 1) * (q - 1) \quad (2)$$

- (4) Choose e to make.
For the RSA algorithm:

$$D(E(M)) = (M^e)^d \bmod n = (M^d)^e \bmod n + E(D(M)) \quad (3)$$

Where M is plaintext. Encryption formula:

$$C = E(M) = M^e \bmod n \quad (4)$$

Decryption formula:

$$M = D(C) = C^d \bmod n \quad (5)$$

2.3.1 Generation of QR Code

The principle of two-dimensional code generation for safety tools is to set a unique two-dimensional code for each safety tool and use the two-dimensional code as the "ID card" of the safety tool to realize the whole process control. The system automatically generates a unique two-dimensional code image based on the warehouse, type, location and other information of the safety tools and appliances, and pastes them on the corresponding safety tools and appliances for downloading or printing. Paste the two-dimensional code of the tool part number and serial number on the shelf where the safety tools are stored, and identify the tool by adding the serial number to the part number. The two-dimensional code of safety tools and instruments should include two-dimensional code of tool number, two-dimensional code of tool part number, and two-dimensional code of tool serial number. The structure of the two-dimensional code is abstracted.

2.3.2 Application of QR Code

- (1) Scan the QR code through the smart terminal to realize the query of safety tools and equipment parameters and the instructions for use of the tools, but the transmission method can only be read through the data or query on the computer, and the query process is relatively cumbersome.

- (2) Scan the QR code through the smart terminal to realize the registration of borrowing and returning of personnel, and the registration of borrowing and returning of tools and instruments. However, the traditional method also requires manual filling in the system, which is relatively cumbersome and complicated.
- (3) Using the QR code label, the borrowing and returning of multiple electrical safety tools of the same model will not be confused. You only need to scan the QR code to know the site of the tool, which is not available in traditional management methods.

3 Realization

3.1 System Management

The user needs to complete identity verification during the use of this system. The specific permissions of the administrator user of the system include assigning role permissions and defining role definitions. The permissions of other users (teams, administrators) are different. When the user opens the homepage, the system judges the user's specific role by checking the cookie. If the user does not store the cookie, the login box will automatically pop out.

The administrator of the system is the main user of the module. Specific operating functions: one is user management, the other is organization management, and the third is basic information management. The user management module can realize user and authority management. Including the addition and deletion of user accounts in the system, the modification of the user's account password, and the respective formulation and modification of the corresponding usage rights in the system in accordance with the specific scope of the user's work.

3.2 Full Life Management of Safety Tools and Equipment

After purchasing in accordance with the regulations of the superior department, the unique QR code will be generated and stored in the warehouse. The corresponding warehouse, category name, type name, type number, scanning of the two-dimensional code of the tool and the warehouse are required. Time, operator, and necessary remarks. After entering the warehouse, the warehouse manager shall go to the tool and appliance warehouse to carry out the operation of requisition and return according to the needs. This part of the realization of the background mainly realizes the data storage operation, that is, the information of the tools and instruments are written into the database.

3.3 Maintenance and Management of Safety Tools and Instruments

According to the maintenance, overhaul, and experiment cycles of different types of safety tools and appliances, automatic tracking is started from the safety tools and appliances in the warehouse, and the maintenance management form of all safety tools and appliances is listed, and a message reminder is sent one month before the maintenance date of the safety tools and appliances. Relevant warehouse management personnel carry out safety tools and equipment maintenance plans, and individually classify and alert the safety tools and equipment that have not been maintained due to expiration.

3.4 Query Statistics of Safety Tools and Instruments

The query function can perform a combined query according to one or more different conditions of purchase time, next test, overhaul time, safety tool number, maintenance time, scrap time, safety tool category name, safety tool identification, and current status, allowing users actual needs are met.

3.5 System Performance Test

Table 1. System network concurrency test results

Concurrency (number of people)	Network communication capability (KBps)
500	3596.17
1000	3181.43
2000	2695.61

As shown in Table 1, according to the current security equipment management business situation, the system's user concurrency is generally maintained at about 400, so the system meets the predetermined requirements in terms of performance.

4 Conclusion

Safe production is the top priority of the power system, and the scientific management of safety tools and equipment can help eliminate potential safety hazards. In this paper, aiming at the information management of safety tools and instruments, with the help of the two-dimensional code technology of the Internet of Things, a set of safety tools and appliances full life cycle management system is designed, and it is applied to the management of safety tools and appliances in the warehouse to improve the management of safety tools and appliances. The level of intelligence has realized information management to ensure that overdue safety tools and equipment will not be used on site to avoid potential safety hazards. The networked management of safety tools and equipment has been realized, and work efficiency has been improved. The system can meet the management needs of power transmission operation and maintenance units for the intelligent and informatization of safety tools and appliances, regulate the management and use of safety tools and appliances, and has good promotion value.

References

1. Ghosh, S.K., Debnath, B., Baidya, R., et al.: Waste electrical and electronic equipment management and basel convention compliance in Brazil, Russia, India, China and South Africa (BRICS) nations. *Waste Manag. Res.* **34**(8), 693–707 (2016)

2. Long, E., Kokke, S., Lundie, D., Shaw, N., Ijomah, W., Kao, C.-C.: Technical solutions to improve global sustainable management of waste electrical and electronic equipment (WEEE) in the EU and China. *J. Remanufacturing* **6**(1), 1–27 (2016). <https://doi.org/10.1186/s13243-015-0023-6>
3. Cui, J., Wang, H., Xue, W., et al.: Application of identity resolution and blockchain technology in the whole industrial chain management of electrical equipment. *J. Phys. Conf. Ser.* **1800**(1), 012011 (7pp) (2021)
4. Preradović, D.M., Mičić, L.S., Barz, C.: Software for improving the quality of project management, a case study: international manufacture of electrical equipment. *IOP Conf. Ser. Mater. Sci. Eng.* **200**(1), 012044 (2017)
5. Zheng, J., He, J., Yu, X.: Discussion on energy consumption management and green development of medical electrical equipment. *Zhongguo yi liao qi xie za zhi = Chin. J. Med. Instrum.* **44**(4):328–330 (2020)
6. Fayustov, A.A., Gureev, P.M.: Electrical and electronic equipment waste management problems. *Ecol. Ind. Russia* **24**(6), 60–66 (2020)
7. Malhotra, R., Mcleod, E., Alzahawi, T.: Management and maintenance of electrical equipment in industrial facilities: procedures for improving safety while saving money. *IEEE Ind. Appl. Mag.* **27**(1), 48–54 (2020)
8. Liu, Q., Wan, X.: Study and practice on distinctive characteristics of mechanical and electrical equipment maintenance and management major in Guangzhou railway polytechnic. *Creat. Educ.* **11**(4), 513–523 (2020)
9. Tsai, W.T.: Recycling waste electrical and electronic equipment (WEEE) and the management of its toxic substances in Taiwan—a case study. *Toxics* **8**(3), 48 (2020)
10. Kitila, A.W., Woldemikael, S.M.: Waste electrical and electronic equipment management in the educational institutions and governmental sector offices of Addis Ababa, Ethiopia. *Waste Manag.* **85**, 30–41 (2019)
11. Jungbluth, N., Keller, R., König, A.: ONE TWO WE—life cycle management in canteens together with suppliers, customers and guests. *Int. J. Life Cycle Assess.* **21**(5), 646–653 (2015). <https://doi.org/10.1007/s11367-015-0982-8>
12. Laso, J., Margallo, M., Celaya, J., et al.: Waste management under a life cycle approach as a tool for a circular economy in the canned anchovy industry. *Waste Manag. Res.* **34**(8), 724–733 (2016)



Building Information Model Monitoring System Based on BIM and Optimized Ant Colony Algorithm

Xiao Cai¹, Lei Pan², Jiawen Gao³, Jinbang Lin², Shilong Zang²,
and Tianhao Zhang^{4,5}(✉)

¹ Nanjing Information Center, Nanjing 210000, Jiangsu, China

² School of Electronic Information Engineering, Nantong Vocational University,
Nantong 226007, Jiangsu, China

³ School of Art and Design, Nantong Vocational University, Nantong 226007, Jiangsu, China

⁴ School of Pharmaceutical and Environmental Engineering, Nantong Vocational University,
Nantong 226007, Jiangsu, China

⁵ Jacob School of Engineering, University of California, San Diego, CA 92093, USA
t6zhang@eng.ucsd.edu

Abstract. Building information modeling is a new technology that has the advantages of three-dimensional visualization and information integration. It can provide a platform for the integration, management and maintenance of different types of information, and realize the exchange and sharing of information. The introduction of BIM technology into building information monitoring can provide a platform for the development of monitoring information analysis and processing functions, and improve the interaction between monitoring information and models. Therefore, this paper designs a building information model monitoring system based on BIM technology, and introduces an ant colony algorithm to optimize system performance. The control variable method was used to test the relationship between the number of building information monitoring, the monitoring frequency and the time-consuming monitoring, and it was concluded that the performance of the system in processing building information can maintain the normal operation of the system.

Keywords: BIM technology · Ant colony algorithm · New building information · Monitoring system

1 Introduction

With the rapid development of my country's construction industry, building information monitoring plays an important role in the inspection of construction quality and safety and the improvement of construction efficiency. Building information monitoring is an important measure to ensure the construction, operation and maintenance of complex buildings, and it has the characteristics of a large number of measuring points, a long monitoring period, a complex monitoring system, and high difficulty in operation and

maintenance. Therefore, the design of building information model monitoring system is of great benefit to the development of information architecture.

Many scholars have conducted research on the building information model monitoring system based on BIM and optimized ant colony algorithm, and the results have been good. For example, a scholar established an online monitoring system based on BIM to monitor the environment of the building structure and display relevant building information data in the BIM software, by drawing monitoring parameter time history curves, structural state graphics, and accessing the monitoring system to collect the original data, vividly show the dynamic status of the monitoring data, and realize the content of custom monitoring [1]. With the help of BIM three-dimensional rendering technology, a scholar established a BIM based visual monitoring system, and used an improved ant colony algorithm to optimize system performance. Safety inspection rules can be integrated into BIM, and automatic hazard identification and prevention during construction can be realized through algorithms Measure planning is to combine the data collected by monitoring to the sensor to realize the direct display of the data collected by the sensor [2, 3]. Although the research results of the building information model monitoring system based on BIM and optimized ant colony algorithm are good, there are few applications of using the system to monitor building information in real life. It is necessary to monitor building information during the construction process to improve the quality of construction safety and efficiency of construction work.

This article describes the characteristics of BIM technology used in building model construction, uses the advantages of BIM technology to build a building information model monitoring system, optimizes the system performance according to the characteristics of the ant colony algorithm to find the best method, and then tests the amount of monitoring information during the operation of the system, the relationship between the monitoring frequency and the time-consuming monitoring function, to verify the feasibility of the system [4].

2 BIM Technical Characteristics and Demand Analysis of Building Information Model Monitoring System

2.1 The Characteristics of BIM Technology Used in Building Model Construction

(1) Visualization

Visualization is what the human eye can see. Applying it to architectural design can play a huge role. In the case of traditional construction drawings, we express the relevant information of building components on the drawings in two-dimensional form. He needs the participants in architectural design and construction to imagine the real form of the architectural shape or concrete components. However, due to the complex and changeable forms and shapes of the existing architectural pursuits, it is difficult to truly show the building itself through the expression of two-dimensional planes. Therefore, the concept of visualization is introduced, and simple plane graphics are displayed in front of people in three-dimensional form; although the concept of three-dimensional renderings has always existed in architectural design, it is produced by understanding the content of related two-dimensional drawings, its

building model does not contain any relevant information of building components, and lacks the linkage between the same components [5]. And what we call BIM visualization means that there is linkage between the same type of components, and the corresponding information can be fed back. The final result can be used to make three-dimensional renderings, and in the entire process of designing building information models, display and model Related communication, discussion and decision-making.

(2) Coordination

Coordination is an important part of the construction from the beginning of the design to the completion of construction. It requires the owner, design unit and construction unit to coordinate and cooperate with each other to prevent problems in the implementation of the project, and to organize all parties involved. Coordinate together to find out the causes and solutions of each problem, and then come up with corresponding remedial measures [6].

(3) Simulation

Most outsiders think that architectural drawings are designed through building models. However, the simulation of BIM technology is not limited to this. In the design process, BIM can simulate the construction model to achieve the goal of design optimization [7]. Such as 4D simulation and 5D simulation. Among them, the 4D model is based on the 3D model with schedule management, and the actual construction is simulated through the construction organization design and schedule plan. The 5D model adds cost management to the 4D model, and compares and analyzes the planned cost and the actual cost [8].

(4) Optimization

From building planning to construction operation is a continuous optimization process, usually not related to BIM applications, but better optimization can be achieved on this basis. Use BIM technology and other optimization tools to optimize the construction and operation of construction projects. According to the complexity of the project, the optimization of the project is realized with the help of advanced technology and equipment, so that the project can reach the most ideal state.

(5) Graphical availability

The drawingability of BIM is reflected in the drawings and materials of integrated pipeline drawings, which is different from the usual CAD drawings. This is the drawing data obtained through the visual display, coordination, simulation and optimization of the information model after the design and construction of the information model.

(6) Integration

The realization of engineering project planning, design, construction and operation management is the core of the integrated management of the entire life cycle of the engineering project. Rely on BIM technology to realize project planning, design, construction and operation management, and then realize integrated management. From the perspective of its integration, the three-dimensional model built by BIM technology is like a large database. In the construction and use engineering, data and information are continuously collected, modified and recorded to achieve integrated management [9].

(7) Parameterization

Parametric BIM technology modeling is modeling through parameter information. It is different from the digital model and the analysis model, the model change can be realized only by adjusting the parameters. At the same time, different parameters can form different graphic elements, and the graphic elements are used as components to construct an overall model through various components.

2.2 Functional Requirements of Building Information Model Monitoring System Database

Requirement analysis is the basis for building a database. Only when the basic work is done fully and accurately can the speed of building monitoring information in the database be improved, the quality of the monitoring information can be guaranteed, and the needs of users can be fully met.

(1) User safety management of monitoring information database

The user security management of the database is to verify the identity of the user who enters the database interface, so as to prevent the information from being arbitrarily tampered with or leaked by non-database users, and to ensure the integrity of the information. In the user login interface designed in this article, when the user name and password input are empty or the user name and password input does not meet the requirements, a corresponding warning window will appear; only when the user name and password meet the specified requirements, will it prompt login success. Only then can you enter the interface of monitoring information management.

(2) Backup and restore of monitoring information database

In order to enhance the security of monitoring information, efficient management of the database is required. If the data in the database is damaged, after backing up the database, you can restore the data to the system and restore the normal operation of the system, thereby avoiding data damage and loss. The database backup interface is to select the data to be backed up from the database drop-down menu, then enter the name of the corresponding backup file, click the browse button to select the backup storage location, and click the backup database button. If the data has been backed up, the program will pop up the word "backed up" [10, 11].

2.3 Mathematical Model Based on Ant Colony Algorithm

When the ant is crawling, it decides its transfer direction according to the amount of information on each path, and realizes the shortest and optimal path selection [12]. In the path search process, the ants calculate the state transition probability according to the heuristic information of the path. $P_{ij}^m(t)$ represents the state transition probability of ant m from city i to city j at time t:

$$P_{ij}^m = \begin{cases} \frac{[\lambda_{ij}(t)]^\alpha \times [\mu_{ij}(t)]^\beta}{\sum_{s \in allowed_k} [\lambda_{is}(t)]^\alpha \times [\mu_{is}(t)]^\beta}, & j \in allowed_k \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Among them, α is the information heuristic factor, which represents the relative importance of the trajectory and the role of the information collected by the ants during the movement. The larger the value, the more likely the ant will choose the path taken by other ants, and the stronger the collaboration between ants; β is the expected heuristic factor, indicating the relative importance of visibility, and also reflects the ants in the process of movement the greater the importance of heuristic information, the closer the probability of state transition is to the existing rules; $\mu_{ij}(t)$ A is the heuristic function, and its expression is as follows:

$$\mu_{ij}(t) = \frac{1}{w_{ij}} \tag{2}$$

In the formula, w_{ij} represents the distance between two adjacent cities. For ant m , the smaller w_{ij} is, the larger $\mu_{ij}(t)$ and $p_{ij}^m(t)$ are. Obviously, the heuristic function expresses the degree of expectation that the ants transfer from city i to city j .

3 Building Information Model Monitoring System Based on BIM and Optimized Ant Colony Algorithm

3.1 Design of Building Information Model Monitoring System

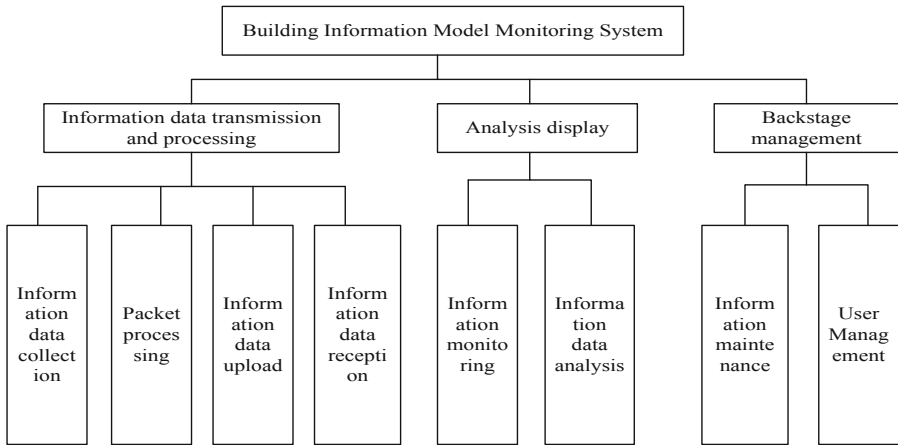


Fig. 1. Building information model monitoring system

Figure 1 shows the design drawing of the building information model monitoring system. The system is divided into three modules, namely, the building information data transmission and processing module, the data analysis display module, and the background management module. There is an information data collection subsystem under the building information data transmission and processing module, which is used to collect building-related information; the data packet processing subsystem is used to store and manage building information; the information data upload subsystem is used

to transmit building information to the database; Information data receiving subsystem, used to receive the building information transmitted by the data upload system. There is a building information monitoring subsystem under the data analysis display module, which is used to monitor the data generated in the process of building construction in real time; the information data analysis system, through the analysis of building data to find problems in the construction process. There is an information maintenance subsystem under the back-end management module to modify building information; the user management subsystem allows users to log in to the building information monitoring system.

3.2 Design of Building Information Monitoring and Early Warning

In the process of real-time monitoring of building information, by comparing the real-time updated monitoring response with the critical value of early warning indicators, if the monitored building information exceeds the critical value, the system automatically alarms and captures abnormal data information, and then realizes the monitoring and warning prompts of the BIM platform include a pop-up warning window to prompt the content of the warning and sound an alarm.

4 Building Information Model Monitoring System Testing

When testing the building information model monitoring system, the monitoring quantity and monitoring frequency are used as variables, and the controlled variable method is used to test the relationship between the monitoring quantity, monitoring frequency and the time-consuming system monitoring.

4.1 Test of the Relationship Between the Quantity of Building Information Monitoring and the Time-Consuming System Monitoring

Table 1. The relationship between monitoring quantity (pieces) and time (seconds)

Number of monitoring	Theoretical time-consuming	Actual time-consuming
1	0.05	0.02
10	1.0	0.1
30	1.0	0.24
80	1.0	0.37
100	1.5	0.45
130	1.5	0.68
170	1.5	0.93
200	2.0	1.12

This test is to control the monitoring frequency to 10 min/time, and the relationship between the amount of building information monitored by the system and the time consumption. It can be seen from Table 1 that the system can process the data uploaded from the information collector in a normal time, that is, the actual monitoring time is less than the theoretical monitoring time. With the increase in the number of building information monitoring, data packets must be monitored in sequence, and the data packets that are not monitored have to wait for the last data packet to be monitored. This increases the waiting time for monitoring and makes the actual monitoring time-consuming. It is also increasing. When the number of monitoring increases, the system should consider increasing the processing memory or increasing the number of packet processing servers to improve the monitoring efficiency of the system.

4.2 The Relationship Between Building Information Monitoring Frequency and System Monitoring Time-Consuming Relationship Test

This test controls the number of building information monitoring to be 50 and 100, and the monitoring frequencies are 10 min/time, 20 min/time, 30 min/time, 40 min/time, 50 min/time, 60 min/time, because the monitoring frequency interval is long, and the results obtained from an experiment have a large error. Therefore, a one-day test is performed to calculate the total time spent in 24 h.

Table 2. The relationship between monitoring frequency (minutes/time) and monitoring time (seconds)

Frequency	50 building information time-consuming	100 building information time-consuming	Multiple relationship
10	43.23	65.66	1.52
20	38.67	57.24	1.48
30	33.29	47.27	1.42
40	27.85	37.88	1.36
50	21.49	28.15	1.31

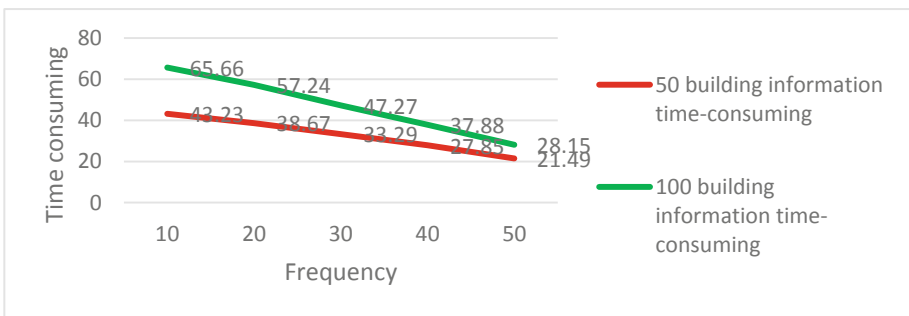


Fig. 2. The relationship between monitoring frequency and monitoring time

It can be seen from Fig. 2 that the shorter the monitoring interval, the more the number of building information detection, and the longer the monitoring time will be. For example, when monitoring once every 10 min, theoretically the time-consuming for 100 pieces of information is twice that of 50 pieces of information, and from the multiple data in Table 2, it can be seen that the multiple is 1.52 times, and there are no multiples of other frequencies. Reached 2 times, indicating that in the same detection interval, that is, when the monitoring frequency is the same, when the amount of building information increases exponentially, the monitoring time will not increase by a corresponding multiple, but will shorten some time consumption. Therefore, in order to ensure the performance of system data processing, an appropriate monitoring frequency should be set to ensure the normal operation of the system.

5 Conclusion

Nowadays, the construction industry has become a pillar industry of my country's national economy. The implementation of construction information monitoring plays a prominent role in the quality management of the construction process and the improvement of construction efficiency. This paper establishes and optimizes the building information model monitoring system based on BIM technology and ant colony algorithm, and conducts experimental tests on the performance of the building information model monitoring system. The test results show that the system performance guarantees the effectiveness of the monitoring results, and the system should be put into the building vigorously. Construction, improve the quality of construction, and stimulate economic growth.

References

1. Kuang, W., Chang, F., Xu, L., et al.: Rail profile identification based on optimized ant colony algorithm. *Zhongguo Tiedao Kexue/China Railway Sci.* **38**(4), 8–14 (2017)
2. Lee, P.C., Zheng, L.L., Lo, T.P., et al.: A risk management system for deep excavation based on BIM-3DGIS framework and optimized grey Verhulst model. *KSCE J. Civil Eng.* **24**(3), 715–726 (2020)
3. Akbar, B.M., et al.: Utilization of design data on conventional system to building information modeling (BIM). *AIP Conf. Proc.* **1903**(1), 1–9 (2017)
4. Cao, D., Heng, L.I., Wang, G.: Impacts of building information modeling (BIM) implementation on design and construction performance: a resource dependence theory perspective. *Front. Eng. Manag.* **4**(001), 20–34 (2017)
5. Sporr, A., Zucker, G., Hofmann, R.: Automated HVAC control creation based on building information modeling (BIM): ventilation system. *IEEE Access* **7**, 74747–74758 (2019)
6. Wang, J., Han, J., Jing, D., Du, P., Wu, C.: MEMS-based strata deformation monitoring system for methane hydrate production areas: a conceptual design. *J. Mar. Sci. Technol.* **23**(3), 425–434 (2017). <https://doi.org/10.1007/s00773-017-0476-4>
7. Won, J.: Quantification of greenhouse gas emissions prevented through building information modeling (BIM)-based design validation - case studies in South Korea. *J. Archit. Instit. Korea Struct. Constr.* **33**(2), 71–77 (2017)

8. Lee, S.Y., Lee, A.Y.: A study on BIPV (building integrated photovoltaic) design process based on the BIM (building information modelling) for architects. *J. Archit. Inst. Korea Plann. Des.* **33**(4), 49–57 (2017)
9. Joon, K.S., Choi, Y.J., Bae, Y.M.: Design and implementation of real-time augmented reality building information system combined with 3D map. *J. Korea Comput. Graph. Soc.* **24**(4), 39–54 (2018)
10. Nkoy, F.L., Hofmann, M.G., Stone, B.L., et al.: Information needs for designing a home monitoring system for children with medical complexity. *Int. J. Med. Inform.* **122**(Feb), 7–12 (2019)
11. Gerrish, T., Ruikar, K., Cook, M., et al.: BIM application to building energy performance visualisation and management: challenges and potential. *Energy Build.* **144**(Jun), 218–228 (2017)
12. Muluken, T., Desbalo, A.A., Hans, I., Bargstadt, J.: Perceived benefits and barriers of building information modeling (BIM) adoption in the AEC sectors of Ethiopia. *Int. J. Eng. Technol.* **9**(10), 777–789 (2020)



Development Status of High-End Equipment Manufacturing Industry Based on Neural Network and Tobit Panel Regression Model

Zilong Di^(✉)

Institute of Economics, Shanghai University, Shanghai, China

Abstract. Equipment manufacturing industry is the cornerstone of a country's industrial development. As its core component, High-end Equipment Manufacturing industry is of great significance to enhance industrial value chain, promote industrial structure optimization and enhance industrial market competitiveness. This paper introduces the Neural Network model and tobit panel regression model, summarizes the HEM industry, the rapid development of China's HEM industry, overseas M & A has become the norm, explores the problems existing in the HEM industry, and studies the innovation and transformation factors of the development level of HEM industry in Heilongjiang Province. The results show that, The TI and transformation of HEM industry in Heilongjiang Province has developed well and is the pillar industry of economic development in Heilongjiang Province.

Keywords: Neural network · Tobit panel regression model · High-end equipment manufacturing industry · Development status

1 Introduction

In recent years, the research of BP Neural Network is highly sought after. Many studies show that the accuracy of the model based on BP NN is higher than that of multiple logistic regression model and multiple probability regression model, and does not need to meet the assumption that the predictive variables are independent of each other. BP NN theory is widely used in enterprises in different industries.

Many experts have studied Neural Network and regression models. For example, Hussain h I, Anwar n, razimi m use generalized regression Neural Network (GRNNs). The results were compared with the traditional panel data regression model by retaining samples. The preliminary results confirm the stability of the data and allow the prediction ability [1]. Noh h, Lee S. Compared the existing methods, such as citation lag distribution, Tobit regression and multilayer NN, to predict future patent citations as an alternative index of patent technology impact [2]. Lee e, Kim D proposed a method of using deep NN (DNN) to detect small traffic lights (TLS) in images taken by vehicle cameras. The TL detector adopts DNN structure, and the encoder decoder has focus regression loss; This loss function reduces the loss of a simple good regression example [3]. Although

the research results of Neural Network and regression model are quite rich, there are still deficiencies in the analysis of the development status of HEM industry based on NN and tobit panel regression model.

In order to study the development of HEM industry based on Neural Network and tobit panel regression model, This paper studies the development of Neural Network, Tobit panel regression model and HEM industry, and finds the radial basis function (RBF) Neural Network. It shows that Neural Network and tobit panel regression model are conducive to the development of HEM industry.

2 Method

2.1 Introduction to Neural Network Model

(1) Introduction of Neural Network Model

Neural Network model is the main method of processing prediction and evaluation in data mining technology. Neural Network model is designed according to the complex network structure of human brain [4]. The human brain can process advanced psychological behaviors such as thinking, emotion and decision-making, but the Neural Network model only imitates this function, that is, the input information is processed and analyzed through the Neural Network model to obtain reasonable output results. The thinking and decision-making function of human brain is based on people's knowledge or previous experience, so the decision-making of human brain is basic and persuasive [5]. Therefore, in order to give full play to the decision-making function of Neural Network model on unknown samples, the learning ability of Neural Network model must be obtained by training known samples. It can be seen that the basis of NN model is the node model of NN, that is, neuron model. The whole NN is composed of each neuron model. Each neuron node has a certain learning ability to process the information obtained by the NN and make the corresponding output decision information. This ability is obtained through NN training [6]. Therefore, NN training is the core function of NN model. In order to obtain an accurate NN decision model, the training of NN must be completed strictly. After network training, NN can realize the functions of prediction and evaluation.

(2) Advantages and disadvantages of Neural Network Model

Advantages of Neural Network model: strong nonlinear mapping ability. Through the construction of multilayer Neural Network Model, the Neural Network model realizes the mapping function from input to output. At present, the most commonly used three-layer Neural Network model is enough to deal with any nonlinear continuous function. This greatly enhances the ability of the model to deal with complex internal relations and the nonlinear mapping ability of the Neural Network model. In the training process, the Neural Network model can obtain the complex relationship between dependent variables and independent variables through autonomous learning, update the error distribution weight of each cycle, and store the learning content in the weight of the model network. The disadvantage of Neural Network model: the model has over fitting phenomenon. According to the algorithm principle of Neural Network model, the Neural Network model is trained

into a model by relying on the original data, and then tested with some data, which reflects the training ability and prediction ability of the model. Generally speaking, there is a positive correlation between training ability and prediction ability. The stronger the training ability, the stronger the prediction ability, the weaker the training ability, the weaker the prediction ability, and the prediction ability will improve with the improvement of training ability. When a certain limit is exceeded, the relationship between training ability and prediction ability will change. With the improvement of training ability, the prediction ability will decline.

2.2 Tobit Panel Regression Model

Tobit model, also known as restricted dependent variable model or sample selection model. When the data to be analyzed has numerical cutting or fragment characteristics, the ordinary least square method is no longer applicable, and the Tobit model using the maximum likelihood estimation method becomes a better choice. Although the model has been widely used in China, there are still some misunderstandings and shortcomings: (1) Tobit model and discrete selection model are confused; (2) It does not distinguish whether the established model is a correlation equation; (3) The economic significance of sample selectivity deviation is not accurately defined. The estimation of Tobit model is mainly divided into three categories: the first is non simultaneous equation model estimation, the second is simultaneous equation model estimation, and the third is panel model estimation [7].

2.3 Development Status of High-End Equipment Manufacturing Industry

(1) Concept of High-end Equipment Manufacturing industry

High end equipment manufacturing industry is a capital intensive and technology intensive industry. Different from the traditional equipment manufacturing industry, it is at the top of the value chain. It is the core link of the industrial chain and has a decisive impact on the industrial competitiveness. High end equipment manufacturing industry provides an important supporting force for modern countries to achieve the goal of industrial transformation and upgrading [8]. At this stage, the development of HEM industry mainly covers five fields, namely aviation equipment industry, rail transit equipment manufacturing industry, intelligent manufacturing equipment industry, satellite manufacturing and application industry, and marine engineering equipment manufacturing industry. Through the industrial upgrading of HEM industry, it can continuously drive the development of China's national economy and other industrial economies. High end equipment manufacturing industry can enhance the competitiveness of the industrial chain, which is the core of the modern industrial system. Today, the development of HEM industry has become the main factor in the competition of powerful countries [9].

(2) The industry has developed rapidly, and overseas M & A has become the norm

In order to promote the development of China's HEM industry, the State Council promulgated the decision on accelerating the cultivation and development of strategic emerging industries in 2010. Since then, China's HEM industry has received

unprecedented attention and achieved rapid development [10]. At this stage, the level of industrial development has been significantly improved, a certain industrial scale has been formed, and the international competitiveness has been continuously enhanced. Some breakthroughs have been made in TI in some subdivided fields: in the field of aviation equipment and satellite applications, it has become one of the few countries that can independently design and build LNG ships, and marine engineering equipment can be alone; In the field of high-speed railway, China has mastered a complete set of high-speed railway technology and become a “large country of high-speed railway technology”. High speed railway technology has begun to go abroad to the world. In the next few years, more and more enterprises will join the army of overseas mergers and acquisitions to obtain nuclear technology. At the same time, it should be noted that different corporate culture, personnel status and capital operation modes will increase the operation risk after M & A. If the reorganization is not completed for a period of time, it will also cause stock market turbulence and increase enterprise operation risk [11].

(3) Problems in High-end Equipment Manufacturing industry

Although the innovation ability of China’s HEM industry is improving day by day, in a considerable part of High-end Equipment Manufacturing fields, key nuclear technologies are still controlled by foreign enterprises. Chinese enterprises can only use these technologies through purchase or introduction, which increases the production cost and investment of products, and will be limited by foreign technology suppliers everywhere. The industrial chain system is imperfect. At this stage, China has not been able to independently produce many related basic components and nuclear key components. Most basic parts and nuclear key parts also need to be imported from abroad. At present, the development of China’s HEM industry also shows the characteristics of regional high convergence. Based on the strong support of the state and the development of key training, in the foreseeable future, China’s HEM industry will enter a period of rapid growth, and its comprehensive strength and international competitiveness will be greatly improved. Initially meet the needs of industrial upgrading and transformation and the cultivation and development of strategic emerging industries, and become one of the world’s major high-end equipment manufacturers. It is worth noting that with the rapid growth of fixed asset investment, the scale and output of construction machinery and machine tool industry rank first in the world, and have reached a record high in China’s equipment manufacturing industry. The technological level of rail transit equipment industry and aircraft manufacturing industry has also made breakthrough progress. In general, China’s High-end Equipment Manufacturing industry has begun to internationalize its strategic layout, and rail transit equipment and other industries have taken the lead in going abroad and participating in international market competition.

2.4 Radial Basis Function (RBF) Neural Network

When the input signal approaches the central range of the basis function, the hidden layer node produces a large output. When the input signal is far from the central range of the basis function, the output of the hidden layer node decreases. The input of hidden layer node is Eq. (1):

$$xh = |c - xi| * bih \quad (1)$$

Where X_i is the hidden layer input vector; C is the center of the input data matrix; BIH is the hidden layer threshold.

It is generally a nonlinear functional relationship, θ is the threshold of neuron, and the output of neuron J can be expressed as Eq. (2):

$$O_j = f\left(\sum_{j=1}^n \varpi_{ij}x_j - \theta\right) \quad i \neq j \quad (2)$$

Among them, ϖ_{ij} is the weight of connection, O_j is the output of neuron, and n is the functional relationship of transfer.

Because the unit proportion of each index is different, it is necessary to convert the dimensional expression of the index into a dimensionless expression. Therefore, according to formula (3):

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

Where, μ mean value and σ standard deviation.

3 Experience

3.1 Object Extraction

This paper adopts the horizontal comparative analysis of the data of three provinces. We can clearly see the ranking of Heilongjiang Province in the three provinces, so as to understand the relative strength of its indicators. Compared with the longitudinal comparative analysis, it is more convincing, and because there is no specific identification scope of HEM industry, it is difficult to obtain the longitudinal data of HEM industry in Heilongjiang Province. Factor analysis was used. As a generalization of principal component analysis, it is an effective method to solve the problem of index information overlap in evaluation, and can achieve good dimensionality reduction. Through the factor analysis of the development level evaluation system of HEM industry, extract the principal component factor, calculate the principal component factor score, and get the development level score of HEM industry in various provinces and cities, which can better understand the development level of HEM industry in various provinces and cities.

3.2 Experimental Analysis

Data envelopment analysis (DEA), analytic hierarchy process, fuzzy comprehensive evaluation, artificial Neural Network model and principal component analysis can be used to evaluate the development level of HEM industry. Data envelopment analysis is a linear programming model. Analytic hierarchy process and fuzzy comprehensive evaluation method are highly subjective. The evaluation of artificial Neural Network model also needs the help of expert scoring or fuzzy comprehensive evaluation.

4 Discussion

4.1 Factor Analysis Results of Development Level of High-End Equipment Manufacturing Industry in Heilongjiang Province

According to the constructed principal component factor score model and comprehensive score model, the principal component factor scores of provinces and cities and the comprehensive scores and ranking of HEM industry are obtained by Excel, as shown in Table 1.

Table 1. Score and ranking of each component factor

Region	Guangdong	Heilongjiang	Hubei
F1	1.272	0.167	0.672
Ranking	1	3	2

It can be seen briefly from the above that the innovation transformation factor of Guangdong Province is 1.272, that of Heilongjiang Province is 0.167, and that of Hubei Province is 0.672. We can see the specific results clearly in Fig. 1 below.

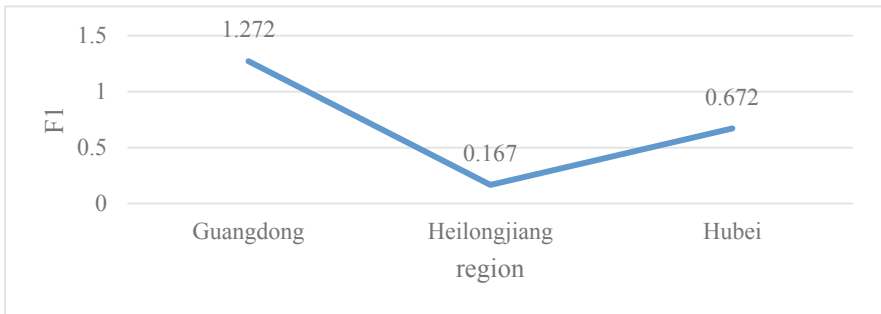


Fig. 1. Score and ranking of each component factor

It can be seen that the top two scores of the development level of HEM industry are Guangdong and Hubei respectively. Heilongjiang Province ranks third among the three provinces and cities, at the lower middle level. The results show that the TI and transformation of HEM industry in Heilongjiang Province has developed well and is the pillar industry of economic development in Heilongjiang Province.

4.2 Regional Distribution of Intelligent Manufacturing Industrial Parks in China

According to the incomplete statistics of the National Bureau of statistics, by 2021, there will be 592 intelligent manufacturing industrial parks under construction or completed in China. Although many small intelligent manufacturing industrial parks have

disappeared, the overall number is still on the rise. The regional distribution of China's Intelligent Manufacturing Industrial Park is shown briefly in Table 2.

Table 2. Regional distribution of intelligent manufacturing industrial parks in China

Region	Midwest	Circum Bohai Sea	Pearl river delta	Changjiang delta	Other
Data	12.5%	15.4%	26.6%	21.7%	23.8%

It can be seen clearly from the above that the number of intelligent manufacturing industrial parks in the central and western regions accounts for 12.5% of the national total, the number of intelligent manufacturing industrial parks around the Bohai Sea accounts for 15.4%, the number of intelligent manufacturing industrial parks in the Pearl River delta accounts for 26.6%, the number of intelligent manufacturing industrial parks in the Yangtze River delta accounts for 21.7% of the national total, and the number of other intelligent manufacturing industrial parks accounts for 23.8% of the national total.

It can also be clearly seen from the above figure that the regional distribution of China's intelligent manufacturing capacity is different. The number of intelligent manufacturing industrial parks in the Pearl River Delta and other regions accounts for more than half of the national total. Overall, the type of Intelligent Manufacturing Industrial Park in China is mainly robot industrial park.

5 Conclusion

The development of HEM industry not only promotes the transformation and upgrading of traditional industries, but also supports the development and growth of other strategic emerging industries. At present, there are still some problems in the development of China's HEM industry, such as low industrial profit margin, lack of independent innovation ability, difficulty in enterprise financing, lack of highly skilled talents and so on. This paper studies the regional distribution of intelligent manufacturing industrial parks in China. The results show that the type of Intelligent Manufacturing Industrial Park in China is mainly robot industrial park.

References

1. Hussain, H.I., Anwar, N., Razimi, M.: A generalised regression neural network model of financing imbalance: Shari'ah compliance as the roadmap for sustainability of capital markets. *J. Intell. Fuzzy Syst.* **39**(4), 5387–5395 (2020)
2. Noh, H., Lee, S.: Forecasting forward patent citations: comparison of citation-lag distribution, tobit regression, and deep learning approaches. *IEEE Trans. Eng. Manag.* **PP**(99), 1–12 (2020)
3. Lee, E., Kim, D.: Accurate traffic light detection using deep neural network with focal regression loss. *Image Vis. Comput.* **87**(Jul), 24–36 (2019)

4. Alizadeh, A., Mosalanezhad, F., Afroozeh, A., Akbari, E., Buntat, Z.: Support vector regression and neural networks analytical models for gas sensor based on molybdenum disulfide. *Microsyst. Technol.* **25**(1), 115–119 (2018). <https://doi.org/10.1007/s00542-018-3942-y>
5. Alikan, E., Sevm, Y.: A comparative study of artificial neural networks and multiple regression analysis for modeling skidding time. *Appl. Ecol. Environ. Res.* **17**(2), 1741–1756 (2019)
6. Park, M., Lee, S., Hwang, S., et al.: Additive ensemble neural networks. *IEEE Access* **8**, 113192–113199 (2020)
7. Sangun, L., Guney, I., Kokcu, P., et al.: Estimation of body weight of *Sparus aurata* with artificial neural network (MLP) and MSP (nonlinear regression)–LR algorithms. *Iran. J. Fish. Sci.* **19**(2), 541–550 (2020)
8. Karamavus, Y., Ozkan, M.: Newborn jaundice determination by reflectance spectroscopy using multiple polynomial regression, neural network, and support vector regression. *Biomed. Signal Process. Control* **51**(May), 253–263 (2019)
9. Petrovi, M.: Assessment of movement of the Belex15 index using linear regression and neural network. *J. Process Manag. New Technol.* **7**(4), 35–45 (2019)
10. Dungore, P.: Analysis of investors’ prediction potential using holt winters model and artificial neural networks. *Appl. Econ. Lett.* **6**, 1–8 (2021)
11. Gazizov, M.R., Grigorian, K.A.: Of neural network model robustness through generating invariant to attributes embeddings. *Russ. Digit. Libr. J.* **23**(6), 1142–1154 (2020)



Construction of Financial Credit Risk Evaluation System Model Based on Analytic Hierarchy Process

Shenghua Qian^(✉)

School of Data Engineering, Tianjin University of Finance and Economics Pearl River College,
Tianjin 301811, China

Abstract. In recent years, affected by the global economic recession, China's economic growth has slowed down, and deep-seated contradictions in the process of economic growth have begun to emerge, and the resulting concentration risks have begun to spread to banks. Effective risk prevention has become an important factor for banks to ensure asset quality. Therefore, the establishment of a complete and effective credit risk assessment system is the key to effective risk prevention. This paper constructs a financial credit risk assessment system, uses the analytic hierarchy method to construct a financial credit risk assessment system model, and then analyzes the actual case of the constructed model, and draws the result. Among the assessment enterprises, the financial credit risk the largest is B for enterprises, the smallest is F. According to relevant data, the construction of this model is effective.

Keywords: Analytic hierarchy process · Financial credit · Risk assessment · System construction

1 Introduction

The government work report pointed out that the factors and conditions that support our country's development have undergone major changes, and contradictions have become increasingly prominent [1, 2], and our country's economy is in a period of structural adjustment and growth changes. In the critical period of overcoming obstacles, the economy is under pressure from enterprises. Soaring increase, thus lending funds to the bank [3, 4]. In the process of banks expanding credit funds, credit risks increase correspondingly, and credit risk management methods have become a necessary condition for banks to operate steadily [5, 6]. Traditional credit risk is actually a default risk. The so-called risk management is a decision-making process used by organizations or individuals to mitigate the negative impact of risks. This process involves goal setting, risk identification and evaluation, risk management plan selection and implementation, and risk management process result monitoring and evaluation [7, 8].

For the study of financial credit risk, some researchers have applied empirical research to the effectiveness of commercial bank credit risk measurement, introduced logit method for empirical analysis, and used neural network technology for comparison, and introduced bank credit risk on the basis of empirical analysis. Model and experimental results show that the logit method has higher prediction accuracy [9]. Some scholars have used discriminant analysis methods and decision tree methods to conduct empirical analysis on the default risk of unlisted SMEs in my country. The results show that both methods can more accurately predict the default risk of a company. It also shows that the decision tree can not only measure a company's default rate, but also analyze the key factors of a company's default, that is, cash flow/total liabilities sum the two ratios of current assets/liabilities are important indicators of valuation [10]. In summary, there are many studies on financial credit risk, but its evaluation system still needs to be studied in depth.

This paper studies the financial credit risk assessment system, analyzes the relevant theories of the credit risk assessment system on the basis of literature data, and then uses the analytic hierarchy process to construct the financial credit risk assessment system, and then uses actual cases to test the model, and draw relevant conclusions through the test results.

2 Research on Financial Credit Risk Assessment System

2.1 Reasons for Financial Credit Risk

First of all, with the acceleration of economic transformation after China entered the WTO, the most fundamental factor affecting credit risk is the structural problem of China's property rights system [11]. The common characteristics of China's existing property rights system and the socialist market-oriented property rights system are as follows: State-owned enterprises and competitive entities are China's main operating entities, not truly independent assets. Private enterprises are not equally protected by the law and cannot be widely recognized by the society. This is the fundamental issue of corporate integrity, and the essence of integrity issues is also a property issue. The prerequisite for integrity activities is that both buyers and sellers have their own property rights [12]. This means that the subject of the transaction should be the land owner or the land owner with independent property rights. The subject has the confidence, ability and motivation to complete the credit loan. Therefore, the issue of China's land property rights system is also affecting the issue of credit.

2.2 Comparison of Evaluation Methods

With the continuous advancement of technology, more and more methods are used for risk assessment. In this article, we mainly choose linear weighted comprehensive evaluation method, analytic hierarchy process, data packaging method, fuzzy comprehensive evaluation method and BP neural network method for comparison. Each method has its own advantages and disadvantages. There is no doubt that there is no perfect way.

When choosing a risk assessment method, the analytic hierarchy process is not the best assessment method, but it is the most appropriate method for the content being assessed. The choice of risk assessment method also depends on the design of the index system and the type of data. The analytic hierarchy process also has some shortcomings, but compared with other methods, it has unique credit risk assessment capabilities. Credit risk is complex, and there may be interactions between different risks. In the business letter risk assessment model, it is not important to understand how different risks individually affect the risk level of the letter of credit. In real-world credit activities, these types of risks often occur at the same time, so only the results need to be obtained.

2.3 AHP Method

(1) Establish a hierarchical model

The purpose of decision-making, the elements to be considered (decision-making criteria), the decision objects are divided into high, medium, and lowest points according to the mutual relationship, which constitutes a hierarchical structure picture. The highest level refers to the main purpose of decision-making and the issues that need to be addressed, the lowest level refers to alternative methods in decision-making, and the middle level refers to the various factors that need to be investigated and the criteria for decision-making. For the adjacent two layers, the top layer is called the target layer, and the bottom layer is called the agent layer.

(2) Create a crisis table (pair comparison).

When deciding the relative weight of different factors at various levels, if there is only a qualitative conclusion, it is usually not easy for others to accept. Therefore, in order to make comparisons, people often use comparison scales to simplify the difficulty of comparing various factors under different properties and increase the accuracy. Therefore, for a certain standard, the shape below it will be a positive comparison and be evaluated according to its key points. a_{ij} refers to the result of comparing the meaning of element i and element j through the nine importance given by Saaty and its assignment. A table composed of the results of the comparison is called a crisis table. Crisis tables usually have the following attributes:

$$a_{ij} = \frac{1}{a_{ji}} \quad (1)$$

(3) Weight calculation method.

The normalized arithmetic mean of the n row vectors of the judgment matrix A is approximated as the weight vector ω , namely

$$\omega_i = \frac{1}{n} \frac{\sum_{j=1}^n a_{ij}}{\sum_{k=1}^n a_{kj}} \quad (2)$$

3 Construction of Financial Credit Risk Assessment System Based on Analytic Hierarchy Process

3.1 Selection of Financial Risk Assessment Indicators

(1) Principles of index selection

- 1) Content. The main body of the financial credit risk assessment system includes multiple entities such as financial companies, third-party logistics companies, financial intermediaries, and core companies. Therefore, when formulating a credit rating model, it is necessary to consider both the lender's own credit status and the influence of other entities on the lender's credit status. Therefore, it is necessary to investigate the influence of industry growth and regional factors on the creditworthiness of the funded company. In short, the financial credit risk assessment system needs to carefully consider various factors that affect credit risk and allow a more complete assessment of the company's credit status based on a thorough investigation.
- 2) Relevance. The construction of the financial credit risk assessment system should start from the entire industry, taking into account the macroeconomic situation, the development characteristics of the industry and the company, and the relationship between the financial company and the parent company. The companies applying for financing should also be considered, and the rating indicator system should be continuously adjusted to make it relevant to the rating model.
- 3) Function. The index selection process should avoid selecting indexes that are inconvenient to collect, difficult to access, and the calculation process is very complicated. Most of the small enterprises in our country are not on the list, the construction of the credit bank for small enterprises has been delayed, and the relevant administrative departments have not supervised them. Some indicators cannot accurately and truly reflect the operating conditions of an enterprise. Therefore, usability and ease of use are important principles for index selection.
- 4) The hierarchical distribution structure is rationalized. According to the financial credit risk assessment system, it is necessary to establish an indicator system from low to high, from simple to complex. Indicators at all levels should highlight the key points as much as possible, and avoid over-complication, over-simplification or absoluteization. The structure should be as logical and efficient as possible, focusing on the true meaning of indicators at all levels.

(2) Index selection

Based on the above analysis, combined with my own theoretical analysis, this article constructs an indicator system from three main aspects:

- 1) Current status of the industry. The factors affecting the company's credit risk status are mainly qualitative indicators such as the domestic macroeconomic environment, industry development, regional development, and past contract performance.

- 2) Apply for the status of the company. This indicator not only considers the financial status of the applicant company, such as profitability, solvency, business capabilities, etc., but also considers the operating and management status and credit status of the loan company.
- 3) Loan assets. In order to obtain credit loans, financial companies mortgage raw materials, stocks and final products to commercial banks. Banks or logistics companies should carefully evaluate the assets of this part of the collateral before granting credit.

3.2 Hierarchical Analysis

(1) Hierarchical structure

Based on the above two first-level indicators, this paper selects 9 s-level indicators such as macroeconomic environment and industrial development and 27 three-level indicators such as macroeconomic situation and legal policy environment to construct the SMEs shown in Table 1. Credit risk identification index system. Figure 1 shows a hierarchical structure diagram, which is an identification indicator system.

Table 1. SME credit risk identification index system

First-level indicator layer	Secondary indicator layer	Three-level indicator layer
Industry status	Macro environment	Macroeconomic conditions X1
		Legal policy environment X2
	Industry growth	Industry development stage X3
		Industry growth X4
		Industry Competitive X5
	Regional factors	Local government support X6
		Marketization situation X7
	Past performance	Default rate X8
Enterprise status	Business management	Manager quality X9
		Staff quality X10
		Corporate Culture X11
		Financial Statement Quality X12
	Profitability	Return on net assets X13
		Return on total assets X14
		Sales profit margin X15

(continued)

Table 1. (continued)

First-level indicator layer	Secondary indicator layer	Three-level indicator layer
	Operational capability	Return on net assets X16
		Return on total assets X17
		Sales profit margin X18
	Solvency	Current ratio X19
		Quick ratio X20
		Asset-liability ratio X21
		Interest protection multiple X22
	Prospects	Net profit growth rate X23
		Sales growth rate X24
		Asset growth rate X25
	Credit history	Bank Credit X26
		Social Credit X27

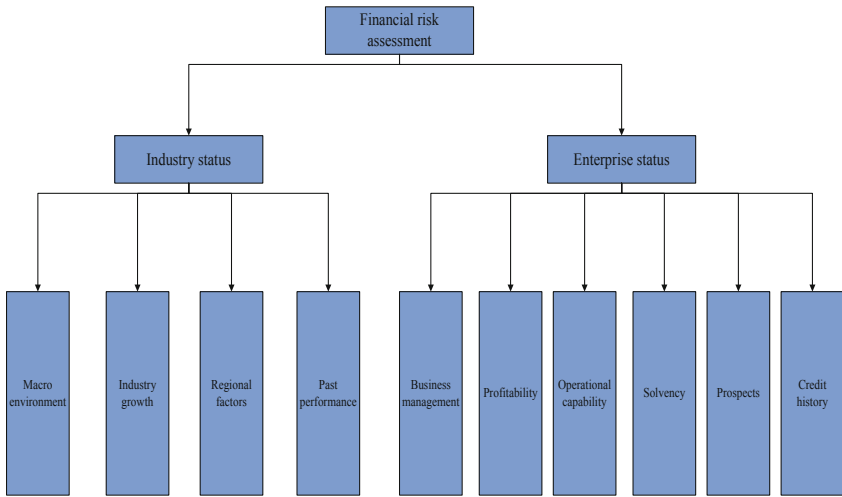


Fig. 1. SME credit risk identification index system

(2) Construction of pairwise judgment matrix and consistency test

In the process of research, this article first determines the scale value of each judgment matrix through the expert scoring method, and then uses the analytic hierarchy process to determine the weight of different indexes.

4 Case Analysis

4.1 Example Data

- (1) It is understood that the parent company A and Bank W have signed a cooperation agreement to provide financial services for the Internet supply chain upstream and downstream of the supply chain where Company A is located. Company B is a supplier of household appliances for specific products. Both Company A and Company B are listed companies in Shanghai and Shenzhen. Company B needs to increase the flow of funds, so it plans to apply for loans from Bank W's online platform by means of supply chain finance. Bank W needs to assess the credit risk of a financial company.
- (2) As a domestic self-operated e-commerce platform, C-core, which has a huge transaction data file of upstream and downstream enterprises, has also signed a cooperation agreement with W Commercial Bank to provide suppliers with supply chain financial services. There are also buyers D Company is active in the field of electrical appliances. This industry is a long-term supplier of Company C. Company D applies for a loan to the bank, and Bank W needs to assess the credit risk of Company D.
- (3) A specific power company E is a listed company, and coal company F is introduced as a supplier in the supply chain. Company F applies to Bank W for a loan, and wants to use the accounts receivable and accounts receivable promised by Company E to obtain a loan. Company E provides a guarantee for Company F, but the company's Internet application level needs to be improved.

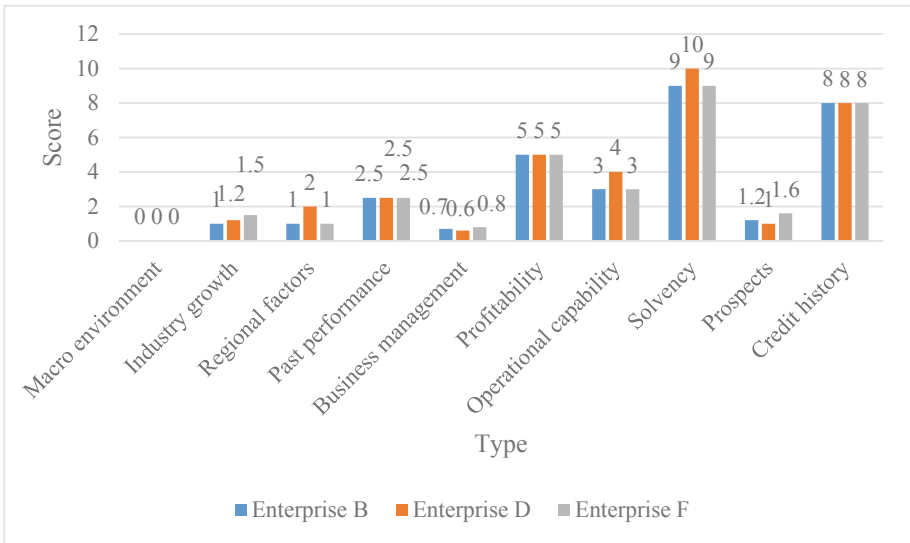
4.2 Scoring Results

According to the above evaluation model, it is first necessary to collect quantitative data and qualitative descriptions of related indicators of related companies from the company's financial statements published on the Internet. Industry and external environmental information can be consulted on relevant magazines and websites. Based on the data obtained, relevant experts and researchers will be required to evaluate the system according to the scoring standards established above. The scoring results are shown in Table 2.

It can be seen from Fig. 2 that the financing risk of enterprise B is greater than that of D, and the financing risk of D is greater than that of F. The credit risk levels are AA, AAA, and A, respectively. In contrast, bank W gives priority to enterprise F in lending, followed by B and D.

Table 2. Scoring result

	Enterprise B	Enterprise D	Enterprise F
Macro environment	0	0	0
Industry growth	1	1.2	1.5
Regional factors	1	2	1
Past performance	2.5	2.5	2.5
Business management	0.7	0.6	0.8
Profitability	5	5	5
Operational capability	3	4	3
Solvency	9	10	9
Prospects	1.2	1.0	1.6
Credit history	8	8	8

**Fig. 2.** Scoring result

5 Conclusions

This paper studies the financial credit risk assessment system model. After understanding the relevant theories, builds the financial credit risk assessment system model based on the analytic hierarchy process, and then uses examples to analyze the results of the experiment. In the assessment of enterprises, Enterprise B's financing risk is greater than D, D's financing risk is greater than F's financing risk.

References

1. Gu, Z.: Evaluation of smart grid construction development impact based on analytic hierarchy process. *J. Phys.: Conf. Ser.* **1176**(6), 62043 (2019)
2. Zha, Y.: Analysis of competitiveness of dry port based on AHP multi-layer fuzzy comprehensive evaluation. *World Sci. Res. J.* **5**(9), 169–177 (2019)
3. Chen, S.F., Charkaborty, G., Li, L.H., et al.: Credit risk assessment using regression model on P2P lending. *Int. J. Appl. Sci. Eng.* **16**(2), 149–157 (2019)
4. Zokirjonov, M.: Methodology of creditmetrics for credit riskassessment. *Int. Finance Acc.* **2018**(2), 7 (2018)
5. Wang, B., Ning, L., Kong, Y.: Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Syst. Appl.* **128**(Aug), 301–315 (2019)
6. Soui, M., Gasmi, I., Smiti, S., et al.: Rule-based credit risk assessment model using multi-objective evolutionary algorithms. *Expert Syst. Appl.* **126**(Jul), 144–157 (2019)
7. Khemakhem, S., Boujelbene, Y., et al.: Support vector machines for credit risk assessment with imbalanced datasets. *Int. J. Data Mining Model. Manag.* **10**(2), 171–187 (2018)
8. Win, S.: What are the possible future research directions for bank’s credit risk assessment research? A systematic review of literature. *IIEEP* **15**(4), 743–759 (2018). <https://doi.org/10.1007/s10368-018-0412-z>
9. Zhang, T., Zhang, W., Xu, W., et al.: Multiple instance learning for credit risk assessment with transaction data. *Knowl.-Based Syst.* **161**(Dec 1), 65–77 (2018)
10. Zhang, Q., Wang, J., Lu, A., et al.: An improved SMO algorithm for financial credit risk assessment – evidence from China’s banking. *Neurocomputing* **272**(Jan 10), 314–325 (2018)
11. Eckert, J., Gatzert, N., Martin, M.: Valuation and risk assessment of participating life insurance in the presence of credit risk. *Insur. Math. Econ.* **71**(Nov), 382–393 (2016)
12. Bandyopadhyay, A.: Managing portfolio credit risk in banks || portfolio assessment of credit risk: default correlation. *Asset Correl. Loss Estimation* (2016). [https://doi.org/10.1017/CBO9781316550915\(6\):235-275](https://doi.org/10.1017/CBO9781316550915(6):235-275)



Product Intelligent Design Model Based on Genetic Optimization Algorithm

Tengjiao Liu¹(✉) and Shalini Kumar Shukla²

¹ Xi'an Fanyi University, Xi'an, Shaanxi, China

² KL University Vijayawada, Vijayawada, India

Abstract. Genetic algorithm is a random search algorithm that uses fitness function as a directional intelligent algorithm. It is based on the evolutionary law of the biological world and has many advantages in practical applications. This paper studies the application of genetic optimization algorithm in product intelligent design, analyzes genetic optimization algorithm and related theories of product intelligent design on the basis of literature data, and then analyzes the product intelligent design model based on genetic optimization algorithm, and test the genetic algorithm cited in the model constructed in this article. The test results show that the average running time of the improved genetic algorithm proposed in the article is reduced. Therefore, under the condition of adopting the elite retention strategy, the excellent individuals generated by the crossover it is retained and not destroyed by mutation, which improves the operation speed of the genetic algorithm, thereby improving the performance of the genetic algorithm.

Keywords: Genetic algorithm · Product intelligence · Intelligent control · Optimization algorithm

1 Introduction

Neural networks have powerful functions and can automatically change the environment [1, 2]. In terms of modeling, neural network is a typical black box learning model. Once the training is completed, the input/output relationship received by the neural network cannot be expressed in an easily acceptable way [3, 4]. Combining the two major advantages of easy expression of fuzzy theoretical knowledge and powerful self-learning ability of neural network, mutual learning can improve the learning and expression ability of product intelligent system [5, 6]. In the field of control engineering, fuzzy neural network is a new type of intelligent calculation method that combines artificial neural network and inference fuzzy logic. As a new technology in the field of artificial intelligence, it is developing in the direction of advanced research and application [7, 8].

Regarding the research and development of product intelligence, some researchers have conducted in-depth analysis of various innovative technologies, based on the principle of combination and function, aiming to optimize the function of the product and realize the innovation of the product as the goal, put forward and explain the concept and innovation the design methods of these methods, and the specific technologies to

implement these methods [9]. Some researchers are considering product functions from the perspective of customer needs and product conceptual design from the perspective of product function design. In addition to analyzing the relationship between product design, conceptual design and functional design, a process model of product design and product conceptual design is established based on the method of functional decomposition, and a network algorithm for this process is proposed [10]. Some scholars have proposed that product design not only combines scientific products, but also combines the relationship between basic parameters and product functions in the design, as well as the functional decomposition tree, and proposes a conceptual design method based on cognitive skills, such as the identification and challenge of important parameters [11]. There is also a multi-level deductive analysis method used by scholars. As a method to guide product life cycle design, QFD can reasonably and effectively transform customer demand information into design requirements at each stage of the product development process, laying the foundation for product functional design [12]. To sum up, there are many research results on product intelligence, but the research on product intelligence control needs in-depth exploration.

This paper studies the product intelligent design based on genetic optimization algorithm, and analyzes the problems in product intelligent design and the related advantages of genetic algorithm based on literature data, and then conducts the product intelligent design model based on genetic optimization algorithm. The built model is built and tested, and draw relevant conclusions through the test results.

2 Research on Product Intelligent Design

2.1 Problems in Product Intelligent Design

The intelligent product controller is composed of a set of rules describing the dynamic behavior of the controller, and is controlled by the empirical knowledge of experts and analog control operators to overcome the influence of non-linear factors. Fuzzy design usually sets basic rules for fuzzy control and membership functions based on fuzzy and inaccurate semantic knowledge or experimental statistical data provided by experts or real operators, and then it simulates human thinking and uses fuzzy reasoning and fuzzy decision-making to achieve control content control to solve actual engineering problems. However, this design often has strong subjective factors. On the one hand, the non-linearity, time-varying characteristics of the controlled object and the existence of random interference lead to some defects and incompleteness in the controller design, especially in the definition of control rules and membership functions, which affects the effect of fuzzy control. In addition, as the number of variables increases, it is difficult to empirically determine the membership function of the multidimensional space t . Effective control rules are selected from a large number of candidate rules, fuzzy rules and membership functions. Automatic repair is also difficult to adjust according to the characteristics of the system, and usually cannot automatically improve the performance of the control system based on accumulated experience.

The input/output relationship of a fuzzy controller is determined by its fuzzy rules and the membership function of fuzzy subsets of fuzzy input and output variables. Therefore, the self-learning, self-adjustment and optimization of fuzzy control can be

realized through the rules of adaptive fuzzy controller and the membership function of fuzzy variables, so that the rules of fuzzy control can be automatically changed, improved and adjusted to achieve the expected control results. So as to continuously improve the control performance in the control process. The rules and functions of the fuzzy controller are determined by their parameters, so the optimization and customization of the fuzzy controller can be transformed into the parameter search optimization problem of fuzzy control rules and member functions.

Due to the large dimension of the parameter space to be optimized, the input and output are indistinguishable and have strong nonlinearity. The traditional gradient optimization method cannot effectively optimize this problem, and the random search algorithm is very complicated and impossible to implement. Genetic algorithm is a stochastic optimization technique. Considering that the problematic solution set needs to be optimized as a group, the optimized objective function can be used to construct the environment and guide the evolution of the solution. The optimization result is better than the traditional optimization algorithm, and the global optimal solution can be approached theoretically.

2.2 Advantages of Genetic Algorithm

- (1) The processing object of genetic algorithm is not the parameter itself, but the person who decodes the parameter set. This function allows genetic algorithms to directly act on the construction of objects.
- (2) Genetic algorithm does not use deterministic rules, but uses probabilistic transition rules to guide the search direction.
- (3) Self-organization, self-adaptation, and self-learning. The genetic algorithm uses the information obtained in the evolution process to organize and search independently. Groups in good condition are more likely to survive and gain genetic structure adapted to the environment.

As a search algorithm, genetic algorithm has the above-mentioned characteristic methods and methods, is convenient to use, and has good robustness. It has proved its unique attraction and benefits in solving and applying various problems, and is widely used in practical problems. At the same time, with the emergence of various mathematical methods and powerful computer simulation tools, genetic algorithms have a wide range of potential applications.

2.3 Genetic Algorithm

This article refers to an improved evolution strategy. The algorithm of this evolution strategy is: first, given the probability of mutation, create an initial population of pop individuals, and use the initial population as the parent, then calculate the applicability, rank the values in descending order of ability price, retain m great men, select and mating creates popular newcomers, merges mating and created newcomers m, maintains and compares the populations, and arranges them in descending order. After the superior individuals are retained as elite individuals, the mutated offspring of mating are produced, and the elite individuals are used to replace the most mutated populations. Poor pop individuals form offspring. If the calculation requirements are met, the calculation will

stop. If the calculation requirements are not met, the offspring population is regarded as the parent population, and the above steps are repeated until the calculation requirements are met.

3 Intelligent Product Design Model Based on Genetic Optimization Algorithm

3.1 Ideas for Intelligent Product Design

The intelligent PID parameter adjustment control methods of the above products have their own shortcomings and limitations. In this article, design a more flexible and smarter product controller, and use feedback to distinguish the output and input signals of the output object. At the same time, through further processing of the error signal, other variables that reflect the characteristics of the measured object can be obtained. Error e , change error ec , and change error change ecc can all reflect some characteristics of the test object. The new controller has no requirements for the exact model to be controlled. Only use errors and error changes to adjust controller parameters in real time to obtain better control results. Figure 1 shows the principle of the smart PID adapter.

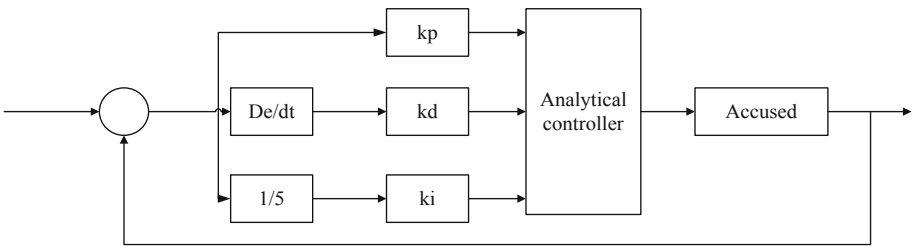


Fig. 1. Principle of Smart PID Adapter

3.2 Intelligent Control Model

In the model designed in this paper, the error forwarding and sending process of BP neural network are carried out separately. The goal is to get the optimal network weights and parameters as soon as possible before starting to optimize the BP neural network. In this process, within the given genetic evolution algebra N , only the promotion process of the BP neural network is carried out. The genetic algorithm is mainly responsible for the evolution of the neural network weights and the execution of the overall search. When a system output error occurs, the deceleration speed starts to slow down. Then start the process of error redistribution of the BP neural network by paying attention to it, and complete the local optimization.

3.3 Control Parameters

(1) Initialization

In the BP neural network, the number of connection weights is very large. For example, take a 3-layer network 4-5-3 (4 nodes in the input layer, 5 nodes in the hidden layer, and 3 nodes in the output layer), a total of 35. In addition, rearranging these parameters into a sequence to form a population string code makes the code string very long, especially when using a binary encoding scheme. In addition, due to the time required for encoding and decoding, the search algorithm of genetic algorithm has become very low. It is necessary to make the encoding more natural to represent the continuous variable optimization problem. It is also convenient to input information related to the specific problem during the optimization process, and the calculation accuracy is not affected by the encoding. The genetic algorithm designed in this paper uses real number coding, because the decoding problem saves a lot of computing resources.

(2) Design of genetic operator

1) Plan a fitness function

The fitness function is used to better guide the evolution of genetic algorithms by determining the suitability of individual populations. The typical design is based on the system output error. The fitness function can be described as Eq. 1.

$$F = \frac{1}{e^2(k)/2} \quad (1)$$

In this evaluation method, the suitability of the population is determined by only one factor. However, in reality, the suitability of the population is affected by a combination of many factors. Therefore, Eq. 1 cannot fully reflect the pros and cons of fitness. Therefore, in some applications of PID control, the fitness function is designed as Eq. 2 to reflect the influence of excess factors on the fitness of the population.

$$F = \frac{1}{e^2(k)/2 + D_k} \quad (2)$$

$e(k)$ is the system output error, and D_k is the rate of change of the absolute value of the system output overshoot.

4 Algorithm Test

This test uses 0–1 coding, single point intersection, and randomly generates the initial population. The overall size is $n = 100$. The calculation result is accurate to 1 decimal place. The mutation probability is $P_{m6} = 0.05$, $P_{m7} = 0.1$. For each test function calculation, the probability of intersection is $P_c = 0.5, 0.6, 0.7, 0.8, 0.9, 1$, and the number of individuals to be processed is $m = 10$. Table 1 shows the GA average execution time test results when the two functions receive different crossover probabilities and reach the end of iteration conditions for the traditional evolution strategy and the improved evolution strategy.

Table 1. Average execution time test results

	0.5	0.6	0.7	0.8	0.9	1
1	4.2	3.9	3.8	3.6	3.3	3.1
	1.6	1.5	1.5	1.3	1.3	1.2
2	30.6	28.2	26.9	25.4	24.5	23.4
	15.7	14.7	14.3	13.9	13.6	12.5

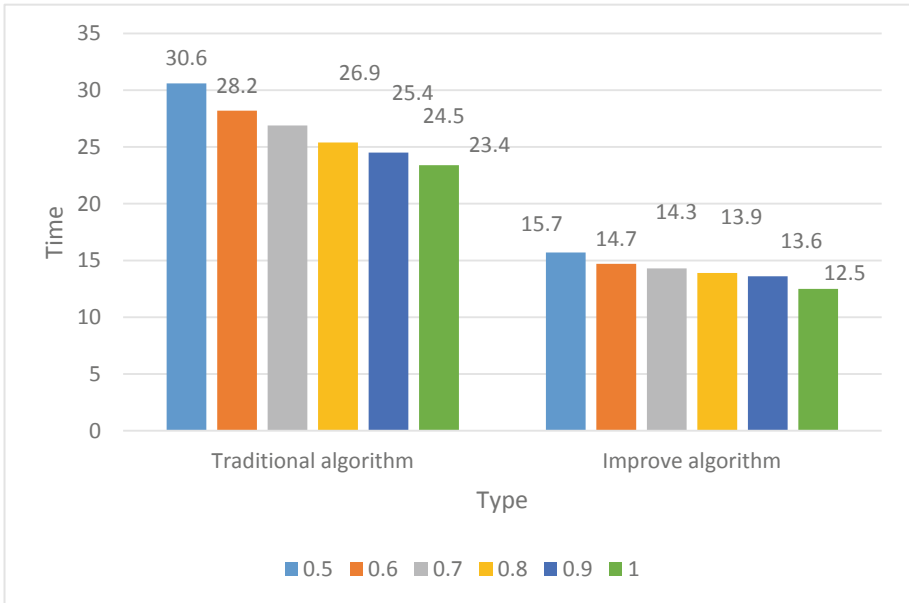
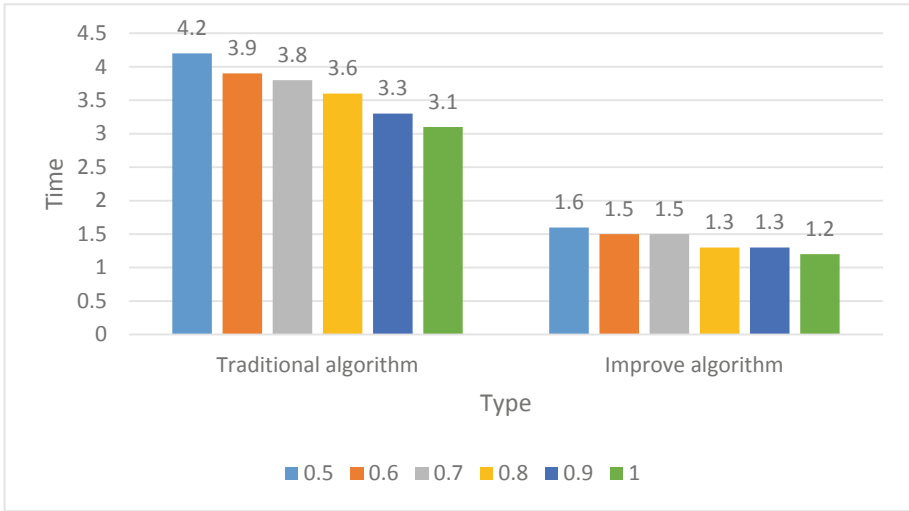


Fig. 2. Average execution time test results

It can be seen from Fig. 2 that for the above two functions, if the probability of intersection is 1, the average running time of each test function is the shortest. As the probability of intersection increases, the average running time gradually decreases. Therefore, the greater the possibility of mating, the greater the possibility of producing better individuals, and the better the performance of genetic algorithms. At the same time, the crossover probability of 1 has been proven to improve the efficiency of genetic algorithms. At the same time, compared with the traditional genetic algorithm, the improved genetic algorithm proposed in this paper reduces the average execution time of each test function when the crossover probability is the same, and the abnormal individuals generated by the crossing keep the interruption unchanged, which speeds up the genetic algorithm. The calculation speed of the algorithm improves the performance of the genetic algorithm.

5 Conclusions

This article focuses on the research of product intelligent design based on genetic optimization algorithm. After understanding the relevant theories, the product intelligent design model based on genetic optimization algorithm is constructed, and the genetic algorithm used in the construction is tested, and the test results are obtained. The optimization strategy proposed in this paper improves the calculation speed of the algorithm.

References

1. Li, P., Ren, Y., Yan, Y., et al.: Intelligent product-gene acquisition method based on K -means clustering and mutual information-based feature selection algorithm. *Artif. Intell. Eng. Des. Anal. Manuf.* **33**(4), 1–15 (2019)
2. Dairou, V., Lageat, T., Delarue, J.: The packaging experience shift: moving from product-centric to life-centric design (2022)
3. Qi, W.: A design exploration of intelligent wearable companion of smartphone for fitness and healthcare. *Int. J. Pattern Recogn. Artif. Intell.* **33**(8), 1959023.1–1959023.14 (2019)
4. Lou, S., Feng, Y., Tian, G., et al.: A cyber-physical system for product conceptual design based on an intelligent psycho-physiological approach. *IEEE Access*, **PP**(99), 1 (2017)
5. Pellicciari, M.: A systematic approach for evaluating the adoption of additive manufacturing in the product design process. *Appl. Sci.* **11** (2021)
6. Tang, D., Zhu, R., Tang, J., et al.: Product design knowledge management based on design structure matrix. *Adv. Eng. Inform.* **24**(2), 159–166 (2018)
7. Roithner, C., Cencic, O., Rechberger, H.: Product design and recyclability: how statistical entropy can form a bridge between these concepts - a case study of a smartphone
8. Xu, T., Zhan, J.: Green product design and pricing decisions in a risk-averse supply chain under alternative power structures. *Math. Probl. Eng.* **2021**(2), 1–20 (2021)
9. Brunner, C.B., Ullrich, S., Jungen, P., et al.: Impact of symbolic product design on brand evaluations. *J. Product Brand Manage.* **25**(3), 307–320 (2016)
10. Li, B.-R., Wang, Y., Wang, K.-S.: A novel method for the evaluation of fashion product design based on data mining. *Adv. Manuf.* **5**(4), 370–376 (2017). <https://doi.org/10.1007/s40436-017-0201-x>

11. Berg, L.P., Vance, J.M.: Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Reality* **21**(1), 1–17 (2016). <https://doi.org/10.1007/s10055-016-0293-9>
12. Tiwari, V., Jain, P.K., Tandon, P.: Product design concept evaluation using rough sets and VIKOR method. *Adv. Eng. Inform.* **30**(1), 16–25 (2016)



Advanced Consumer Psychology Based on Decision Tree Data Mining Algorithm

Yue Lu^{1,2}(✉) and Jie Zhou¹

¹ Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

² Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract. With the advancement of technology in this information age and the rapid development of e-commerce, payment methods have become more convenient and consumer behavior has also undergone tremendous changes. At the same time, the improvement of the quality of life also makes consumers pursue branding and gentrification of consumption. The phenomenon of pre-consumption is becoming more and more common, reflecting the status quo of irrational consumption. Based on the decision tree data mining algorithm, this paper analyzes the impact of monthly income on the number of advanced consumption. It is found that the higher the monthly income, the more the number of advanced consumption. At the same time, it also analyzes which type of advanced consumption is used by groups under 35 and over 35. The number of people is larger, and it is found that 32.86% are using Huabei, 27.21% are using credit cards, 26.50% are online borrowing, and 13.43% are borrowing from friends.

Keywords: Decision tree · Data mining algorithm · Advanced consumption · Consumer psychology

1 Introduction

With the improvement of living standards and the increasing number of loan methods, more and more consumers meet their own commodity needs through advanced consumption. The changes in people's lifestyles and consumption concepts have greatly increased people's expectations for the quality of life. The diversification of consumer decision-making and the development of market economy have led to an increase in irrational factors in people's consumer behavior. Therefore, by studying consumer psychology factors, we can understand the characteristics of people's advanced consumption behavior and help people establish a good consumption outlook.

Many scholars at home and abroad have conducted research on advanced consumer psychology analysis based on decision tree data mining algorithms, and have obtained corresponding research results. For example, a research team expressed insights on the issue of advanced consumption. They believe that advanced consumption is mainly manifested as excessive consumption, emphasizing the surface value of goods and ignoring

the actual application value, just to satisfy people's desire for consumption. Excessive consumption inevitably needs to rely on excessive production to ensure that, and excessive production will inevitably lead to the greedy plunder of natural resources, and then destroy the ecological balance and harmonious development. Advance consumption is separated from the essence and real needs of human beings [1]. A scholar analyzed the consumption needs of people shopping online in the era of information technology, using data mining technology to find the correlation of people's shopping characteristics, and believed that when people are shopping, different consumer psychology ultimately leads to consumption patterns and consumption expenditures. The difference [2]. Although the research results of advanced consumer psychology based on decision tree data mining algorithms are good, advanced consumption will cause economic burdens on people, and advanced consumer behavior should be reduced.

This paper introduces several decision tree algorithms, puts forward the related concepts of advanced consumption and analyzes the pros and cons. According to the number of people using different consumption methods, it is concluded that the income situation is proportional to the number of advanced consumption, and the number of people using Huabei is the largest.

2 Decision Tree Algorithm Classification and Pros and Cons Analysis of Advanced Consumption

2.1 Common Decision Tree Algorithms

(1) ID3 algorithm

The ID3 algorithm is used for predictive classification, and its core is "information entropy", and the information gain comes from entropy calculation as a measure of the discriminative power of divided attributes [3]. The algorithm theory is easy to understand, and the data processing speed in the calculation process is very fast. However, no operation can be performed when the data attribute is missing.

(2) C4.5 algorithm

The C4.5 algorithm is an improved algorithm proposed on the basic properties of the ID3 algorithm to make up for the deficiencies of the ID3 algorithm. The improvement of C4.5 is mainly concentrated in the following aspects: on the one hand, the information gain ratio is used as a measure of resolution, which reduces the possibility of overshooting defects in the ID3 algorithm; on the other hand, it can efficiently process data and continuous information lacking, type data [4, 5]. The calculation classification rules created by the C4.5 algorithm are convenient for people to understand, and the algorithm has a higher accuracy rate, but it is still in the category of "information entropy". C4.5 also has obvious shortcomings, that is, when creating a decision tree, the data training package needs to be reordered many times, and the calculation performance is relatively low. Moreover, the C4.5 algorithm cannot handle data with too much information [6].

(3) SLIQ algorithm

In order to overcome the insufficiency of decision tree algorithm that can only process memory data sets, the proposed SLIQ algorithm can process data with high scalability, fast calculation speed and relatively large scale. The algorithm can run in parallel on multiple processors, greatly improving the efficiency of processing data. The idea of the SLIQ algorithm is to use the method of quickly solving partial sets to determine the splitting conditions of discrete features, and the processing of continuous features uses a combination of pre-arrangement and width technology to build a decision tree. The advantages of this algorithm are simple operation, fast calculation speed, and high prediction accuracy when processing a large amount of data information, which results in higher calculation accuracy when processing training samples with larger data packets. But it is also because the processed data packet greatly affects the memory capacity, causing the algorithm to run lag, and the algorithm's scalability performance is poor [7].

(4) RainForest algorithm

The main idea of the algorithm is to maximize the use of memory resources and create decision trees in permanent memory as little as possible. The core of resource usage is that in any calculation process, the size of the data set is reasonably adapted to the available memory space, and the function of memory resources is played to the greatest advantage.

2.2 The Definition and Pros and Cons of Advanced Consumption

Advanced consumption means that the current income level is not enough to buy the products or services needed now, and consumption is made in the form of installment payment and advance payment. It is a blind consumption behavior for the purpose of enjoyment [8].

Advanced consumption is the actual internal demand of consumer groups to drive social and economic development, and it is also the change in consumption concepts caused by my country's entry into a moderately prosperous society. my country has promulgated a variety of consumption policies to mobilize consumers' enthusiasm for purchasing goods. Credit consumption, mortgage loans, and personal loans all guide people to consume ahead of time, which inspires more consumers to dare to use tomorrow's money to pay today. In fact, the lending behavior of the government and banks also reflects the concept of advanced consumption. These lending measures have promoted the development of local construction and business. Similarly, the change in people's consumption concepts has played an important role in expanding domestic demand for products and promoting social consumption. But advanced consumption can also bring negative effects. Since the pursuit of advanced consumption is material enjoyment, excessive advanced consumption will make society's money and cultural ideology more serious. Some unrealistic commodity demand not only wastes consumers' money, but also wastes producers' resources. If consumers make advance consumption through loans and cannot repay the funds when they are due, it will affect the reputation of consumers and cause a crisis of trust [9, 10].

2.3 Characteristics of Advanced Consumer Psychology

(1) Consumption impulse psychology

Many consumers make consumer decisions that are affected by consumer impulse factors, such as commodity price cuts and consumer sentiment. In the process of consumption, people will also consider economic strength, whether they are practical or not, but the impulse to consume will make people ignore the actual value of consumer goods, especially some luxury goods and high-end goods. Preference will generate a strong desire for consumption, leading people to advance future assets to meet existing consumption needs [11].

(2) Personalized consumer psychology

With the development of today's consumer goods market, consumer products are becoming more abundant, and consumers can purchase goods or services according to their personal psychological needs. Today's consumer groups are more inclined to innovate and change, have a strong curiosity, and pursue personalized consumption. What they choose is not only the practical value of the commodity, but also individual differentiation, and personalized consumption has become the mainstream consumption of the new generation [12].

(3) Comparing psychology

Because people live in a diverse living environment, the consumption behaviors of the surrounding people will affect their own consumption concepts, especially when people's mentality is strong, if their classmates, colleagues, and friends purchase luxury goods, they think they cannot be compared by them, you will have a psychology of comparison, leading to impulsive consumption, buying some things you don't need, or even overdrawing your living expenses.

(4) In the mind of the consumption plan

People have a certain degree of planning for advanced consumption. Since it is advanced consumption, it means that current income cannot meet the demand for commodity prices. People must have sufficient economic strength to fill the current income gap and balance the income brought by advanced consumption. Debts and own income situation. Generally speaking, people have certain cognitive abilities and have plans for their own advanced consumption behavior.

2.4 ID3 Algorithm

If a database is neat and orderly, that is, there are no heterogeneous tuples, there is no need to continue the division. In addition, every time the algorithm divides the subset, it selects the split state that maximizes the order. Information gain is selected as a measure of the discriminative ability of partitioning attributes. The so-called information gain is the difference between the information needed to correctly classify a set of data and the information after the split. The specific calculation method is: calculate the entropy of the original data set and the difference between the weighted sum of entropy of the data set after the split, and the weight is represented by the proportion of the subset in the total data set after the split.

Let A be a set containing a data samples and n different category attributes Y_i , where $i = 1, 2, \dots, n$. Therefore, if the number of samples in the category Y_i is A_i , the amount of information required to classify the sample set A is defined as:

$$C(a_1, a_2, \dots, a_n) = - \sum_{i=1}^n t_i \log_2 t_i \quad (1)$$

Among them, t_i represents the proportion of the number of samples A_i in the category Y_i to the total samples, and the calculation formula is:

$$t_i = \frac{a_i}{A} \quad (2)$$

3 Research on Advanced Consumer Psychology Based on Decision Tree Data Mining Algorithm

3.1 Research Purpose

Under the situation of rapid economic development in my country, advanced consumer psychology and consumer behavior can fully mobilize the enthusiasm of individual consumption for innovation, and can also drive domestic demand for economic development, so as to maximize enterprise production and promote producers to create social and economic value. But at the same time, it will also cause individual economic burdens. Therefore, analyzing people's advanced consumption behavior under the way of many pre-expendable future assets will help people form a reasonable consumption concept and avoid blind consumption behavior.

3.2 Research Methods

This article interviews the monthly income and advanced consumption patterns of 283 employees of a company, and uses this information to explore the impact of monthly income on advanced consumption and the usage of different advanced consumption patterns in different age groups.

4 Analysis of Advanced Consumer Psychology Based on Decision Tree Data Mining Algorithm

4.1 The Impact of Income on Advanced Consumption

Figure 1 shows the average number of people's advanced consumption in the next month under different income conditions. The number of advanced consumption with a monthly income of less than 2,000 yuan is 2 times, and the number of advance consumption with monthly income of 2,000–4,000 yuan is 3 times. The number of advance consumption between 4000–6000 yuan is 5.5 times, the number of advance consumption with a monthly income of 6000–8000 yuan is 7 times, the number of advance consumption

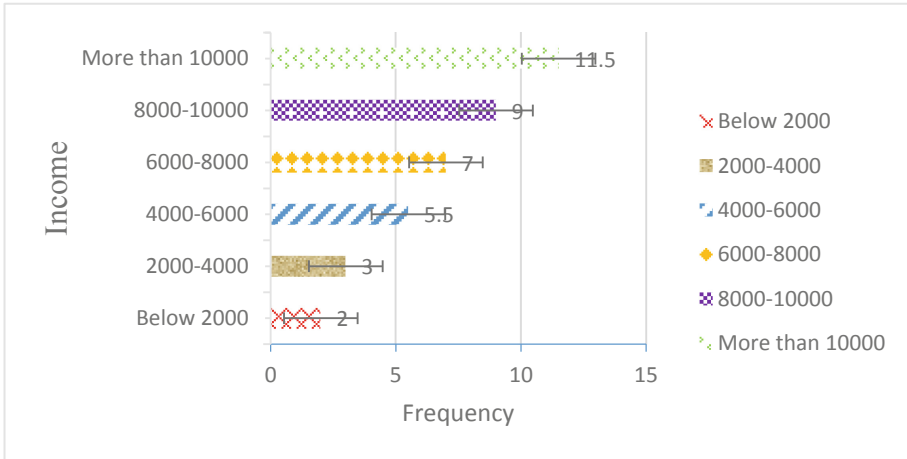


Fig. 1. The impact of income on advanced consumption

with a monthly income of 8000–10000 yuan is 9 times, and the monthly income is more than 10000 yuan, the number of advance purchases is 11.5. According to the above analysis, with the increase of monthly income, people’s number of advanced consumption will increase, that is to say, income is directly proportional to the number of advanced consumption.

4.2 Analysis of Advanced Consumption Patterns

Table 1. Advance consumption method

	Under 35years old	Over 35years old	Proportion (%)
Huabei	56	37	32.86
Credit card	30	47	27.21
Online loan	42	33	26.50
Borrow from friends	25	13	13.43

As shown in Table 1 and Fig. 2, the interviewees are divided into two age groups, those under 35 and over 35. Among the people who use Huabei to spend, there are 56 people under the age of 35, 37 people over the age of 35, accounting for 32.86 of the total number of people; among the people using credit card spending, 30 people are under 35 years old, and those over 35 years old. There are 47 people, accounting for 27.21% of the total; among the people who use online loans to advance consumption, 42 people are under the age of 35, 33 people over the age of 35, accounting for 26.50% of the total; among those who borrow money from friends for consumption, 25 are under the age of 35, and 13 are over the age of 35, accounting for 13.43% of the total. Based on

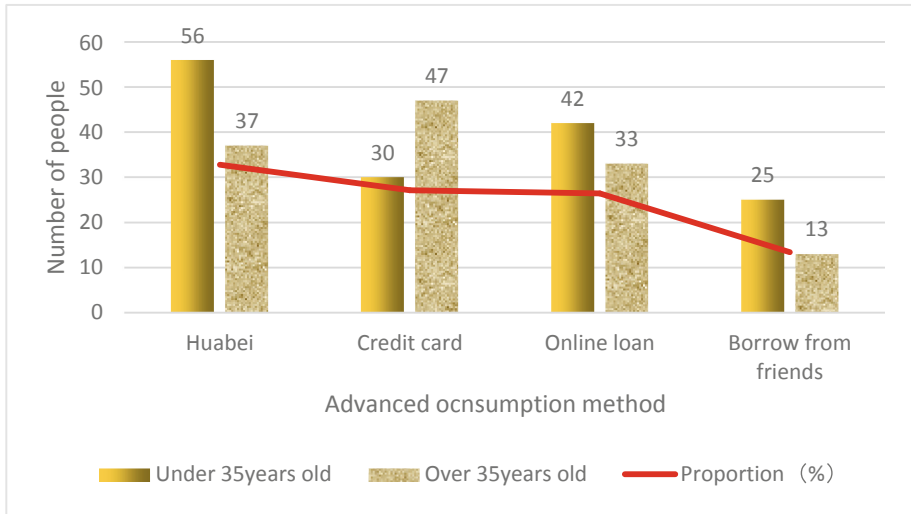


Fig. 2. Advanced consumption patterns by age

comprehensive analysis, the number of people who use Huabei for pre-consumption is the largest, and the number of people who borrow money for consumption is the smallest. However, because pre-consumption expenditure exceeds income expectations, people will incur debts. It is also necessary to minimize the use of online loans, credit cards and other methods to consume large amounts. Therefore, it is necessary to consume rationally to avoid reducing the quality of life due to excessive expenditures, and to prevent imbalances in income and expenditure.

5 Conclusion

People's consumption concepts are changing with the development of society and the progress of the times, and advanced consumption has become a popular way of consumption. Exceeding consumption is a normal consumption behavior of people, but we do not advocate the consumption habit of often paying for current consumer goods through advances in future assets. In the principle of daily consumption, we must clarify the consumption positioning, correct the consumption mentality, and cannot follow the trend, resulting in a comparison and vanity psychology. Only in this way can a healthy consumer psychology be formed and economic developmental consumption can be promoted.

References

1. Harman, B.: Routledge international handbook of consumer psychology. Qual. Market Res. **23**(2), 329–332 (2020)
2. Guha, M.: The dark side of social media: a consumer psychology perspective. Int. J. Soc. Psychiatry **64**(5), 513–515 (2018)

3. Kumar, P., Polonsky, M.J.: An analysis of the green consumer domain within sustainability research: 1975 to 2014. *Australasian Marketing J. (AMJ)* **25**(2), 85–96 (2017)
4. Sett, R.K.: Inflexible microcredit contracts and their discontents: a theoretical perspective based on consumer psychology-ScienceDirect. *IIMB Manag. Rev.* **32**(1), 96–107 (2020)
5. Rao, V.R., et al.: Emerging trends in product bundling: investigating consumer choice and firm behavior. *Cust. Needs Solut.* **5**(1–2), 107–120 (2017). <https://doi.org/10.1007/s40547-017-0075-x>
6. Hota, M., Bartsch, F.: Consumer socialization in childhood and adolescence: impact of psychological development and family structure. *J. Business Res.* **105**, 11–20 (2019)
7. Wu, L., Shen, H., Li, M., et al.: Sharing information now vs later the effect of temporal contiguity cue and power on consumer response toward online reviews. *Int. J. Contemp. Hosp. Manag.* **29**(2), 648–668 (2017)
8. Wiebe, J., Basil, D.Z., Runté, M.: Psychological distance and perceived consumer effectiveness in a cause-related marketing context. *Int. Rev. Public Nonprofit Marketing* **14**(2), 197–215 (2016). <https://doi.org/10.1007/s12208-016-0170-y>
9. Kuchmaner, C.A., Wiggins, J., Grimm, P.E.: The role of network embeddedness and psychological ownership in consumer responses to brand transgressions. *J. Interact. Marketing* **47**, 129–143 (2019)
10. Shi, Y., Chen, W., Zhu, Y.: Study on prediction model of number of rainstorm days in summer based on C5.0 decision tree algorithm. *Meteorol. Environ. Res.* **10**(02), 60–64 (2019)
11. Amarendra, C., Reddy, K.H.: Performance analysis intelligent-based advanced PSO algorithm and testing of real-time matrix converter electrical system. *Soft. Comput.* **24**(18), 14209–14220 (2020). <https://doi.org/10.1007/s00500-020-04789-9>
12. Zhao, Y.: Research on personal credit evaluation of internet finance based on blockchain and decision tree algorithm. *EURASIP J. Wirel. Commun. Netw.* **2020**(1), 1–12 (2020). <https://doi.org/10.1186/s13638-020-01819-w>



Design and Implementation of Hotel Management Information System Based on Intelligent Terminal

Yanjie Yu^(✉)

Liaoning Vocational College of Light Industry, Dalian 116100, Liaoning, China

Abstract. With the rapid development of modern computers and emerging media and network information technology, the hotel industry is becoming more and more information-based management. Therefore, in the information environment, the hotel information system is effectively designed and implemented to help consumers complete online hotel reservations more conveniently, which has become a key issue in the development of China's hotel industry. This paper aims to study the design and implementation of hotel management information system based on smart terminals. Based on the analysis of the characteristics of the hotel management information system and system performance requirements, the hotel management information system is designed and the performance of the system is tested. The test results show that the system response time is within 3 s, and the CPU occupancy rate of the system modules is also within 30%, which meets the performance requirements of this article.

Keywords: Smart terminal · Hotel management information system · Guest room management · Reservation management

1 Introduction

Nowadays, many fields of social science have entered the information age, and the results have also reached the forefront of each century. The management information system has exerted a huge influence in the development of human economy and society, and the wide application of informatization has also exerted a great influence on many aspects of human economic and social life [1, 2]. One of the important signs of our country's modernization is the scale of development and the widespread use of management information systems.

In the rapidly changing environment of today's information age, informatization is also a new wave of worldwide development. Interdisciplinary enterprise information management systems have also become the focus of attention of the entire society, thereby promoting enterprise informatization and management in all walks of life. Many developed countries abroad are also one step ahead in this aspect of management and operation, and have used advanced computer technology and perfect development models [3, 4]. In China, the hotel industry has gradually begun to implement an information

management model. The information management system has been widely used in many aspects of China's hotel industry management. It has always been the first in the industry due to its high economic and social benefits [5, 6]. Our country's hotel industry has had a special history of development so far. In the context of the increasingly rapid development of information science and technology, with the improvement of the hotel information system in our country, it is gradually developing towards differentiation. Now many hotel companies can choose the most ideal hotel management information system according to the actual situation [7, 8]. However, in developing countries like China, although small hotels occupy an absolute proportion of the city, they are often unable to adopt the most ideal hotel management software due to the constraints of talents and funds, and now they still use the original management methods. It's easy to make mistakes, and its operating efficiency is poor, and it can't keep up with the pace of social development [9, 10]. Nowadays, hotel management software developed by developers has more complete functions, but the price is also very high. Some system functions are no longer available in small hotels. At present, small hotel systems urgently need to be reasonably priced, easy to use, impeccable in business, and meet management requirements in business to meet the needs of enterprises to expand and develop related hotel system functions.

On the basis of consulting a large number of relevant references, this paper designs a hotel management information system based on the characteristics of the hotel management information system and system performance requirements. The system mainly includes four modules, namely user module, guest room management module, reservation module and message management module. After implementing these four modules, the performance of the system is tested to verify whether the system meets the requirements of this article.

2 Design and Implementation of Hotel Management Information System Based on Intelligent Terminal

2.1 Features of Hotel Management Information System

(1) Auxiliary system

The hotel management system is mainly people-oriented, and the computer is only an important tool for information processing. Managers make administrative decisions through computers that achieve efficient operation and management goals, and the results of data processing information. The hotel's information management system is a brand-new man-machine management system. Through the high-speed information processing of the network system, hotel managers can quickly grasp the required information [11, 12]. The use of auxiliary management can make the hotel management more scientific, and the management of corporate operations is more standardized, standardized, and refined.

(2) Open system

The hotel information system is an open system with input/output functions that can enter all accounts, registration documents, accounts, reports, etc. It can also export all statistical reports, summary tables, etc., to manage the logistics and capital flow

of the entire hotel in the entire production process. Not only can it analyze and adapt to the environment, but it can even improve it to a certain extent. Therefore, it is possible to improve hotel marketing activities through systematic sales planning.

(3) Feedback system

As a complete control system, the system handles certain business data for the hotel. Because hotel operations are a constantly changing environment, unless they are adjusted and expanded during the operation process, they will often lose their effectiveness because they are far from reality. Therefore, in order to prolong the life cycle of the system, it is necessary to customize the internal processing method according to the output information and external information, and expand the corresponding processing function at any time. This feedback function makes the output result information more accurate and practical.

(4) Hierarchical system

Modern hotel management has a clear hierarchy system. That is to say, the functions of managers at all levels are clear, divided into tools, and subordinates obey their superiors. In order to meet this kind of management, the hotel information system company specially designed the corresponding level of management software. This level usually includes the following three levels: the lowest level of operation level. The main purpose is to record and manage some basic data, and the main purpose of using the hotel information system is to improve operating efficiency and service level; the middle level is the management level, whose main function is to manage comprehensive data, and the main purpose is to enhance the accuracy of management. It makes the planning, organization, management and incentive work of the entire hotel more efficient and targeted; the highest level is the decision-making level, which is used by high-level guest room managers, who make management decisions based on the information extracted from the hotel management information system. In order to improve management efficiency, this is the real value of computers in hotel application, cost control decision-making, financial planning decision-making, goal setting, etc. At that time, the implementation of the hotel management information system will enter an advanced stage, that is, the hotel decision support system stage.

2.2 System Performance Requirements

(1) Reliability

Reliability requirements are the most important operating requirements of the entire system. Because the reliability of the hotel management system is very important, it is necessary to strengthen the database construction and improve the security of information backup. The system implementation process must ensure personalization at the application level. In short, you cannot change the database backup data at will, regardless of whether the operating system is faulty or not. This is a principle that must be followed in system development, but it is also a common problem in system development. It is not difficult to understand that reliability is the main non-functional requirement for the research and development of hotel management information systems.

(2) Scalability

The operation and management of the hotel is not static, but depends on the market economy and the external environment. This stipulates that the hotel management information system must have strong scalability. Managers can analyze the relevant modules based on the actual development of the hotel. In order to avoid the impact of adding new modules on the operation of the entire system, the module architecture developed according to the design requirements is strictly controlled, and a flexible and reasonable design is carried out in accordance with the specific actual situation. It can effectively reduce the difficulty of replacing system modules later.

(3) Security

A key indicator to effectively evaluate the service quality of an information system is the reliability of the information system. Although the hotel information management system does not have a networking function, the information system includes financial information, customer order data, and service personnel information system data. Once there is deliberately damaged system information, it will cause the entire system to be damaged, signal disorder, and unexpected losses to the hotel. Therefore, the safety and reliability of the entire system must be guaranteed. In terms of security, the system must not be used beyond the scope of the customer's responsibility, or work data may be damaged, damaged, or leaked, and appropriate protective measures must be taken. When considering the security aspects of the operating system, you should also consider the security of the system itself.

3 Experiment

3.1 User Module

System user interface requirements are becoming more and more artistic, and interface user requirements are becoming more and more complex. The style of the interface is very important to the user experience of the system. With the development of computer science, psychology, knowledge and design, there are more and more intersections between disciplines. Interface design quality, as a means of exchanging and transmitting information between humans and computers, has become an important indicator of software usability and an important measure of system quality.

The user login module is the prerequisite for the user to log in to the system. Before logging in, the user must confirm the user ID, and the user must enter the user name and password for authentication. If the user name and password are correct, enter the system; otherwise, you must enter it again. There are only three opportunities to re-enter. The realization of the user interface module is mainly to input the corresponding data through the controls on the page. CLogin() runs in the background, compares the database user name and password, and then compares the results in the next step. After receiving the user name and password, compare them in the database to determine whether the connection is successful. Throughout the process, the password uses MD5 encryption technology to protect the system. The logic function used by MD5 encryption technology is shown in the following formulas 1 and 2:

$$F(x, y, z) = (x \& y) / ((\sim x) \& z) \quad (1)$$

$$G(x, y, z) = (x \& Z) / (y \& (\sim z)) \quad (2)$$

In the formula, & and ~ respectively represent AND and NOT operations.

3.2 Guest Room Management Module

(1) Room information registration

Many factors, such as the expansion and scale of the hotel, require hotel managers to add new rooms to the normal operation of the hotel. Enter the new guest room section, enter the basic information of the new guest room as prompted on the page, and click the submit button. The system first verifies the format of the information sent by the administrator, and if the verification is correct, it performs a second verification, such as whether the sent guest room numbers have the same record. Finally, send the database. If the administrator prepares to re-enter before submitting, he can click the reset button to return the completed data to the page.

(2) Maintenance of guest room information

On the guest room information management page, click the change button, the system runs the query function, and the query result is linked to the change page. When the user changes the room information and clicks the "Save" button. The room management page has a picture with a logo, allowing to change or delete the room information. Select the delete icon to delete the information corresponding to the current room.

(3) Guest room status management

The database table where the guest room status is linked to the page function is the guest room information table. When the user clicks on the guest room status page, the system will read guest room information such as the guest room number and current status. Then you will receive the room status as a crisis situation. The value "0" indicates that the room is vacant, and the value "1" indicates that the room is occupied.

3.3 Reservation Module

(1) Check-in and check-out date selection module

When booking a hotel, the user first enters the check-in and check-out time, queries the database according to this time period, returns the relevant available hotel room resources, and returns the available rooms on the room selection page and the page. All that needs to be called are dynamically created and displayed. Reserved users cannot view the resources of these unavailable rooms. The flowchart is shown in Fig. 1.

In order to unify the format of the imported date, use the calendar control to select the date instead of manually importing it. Since this is an open booking

model, the calendar check should set the target date to a maximum of 90 days after the current date to avoid management confusion due to too many family versions released in the future.

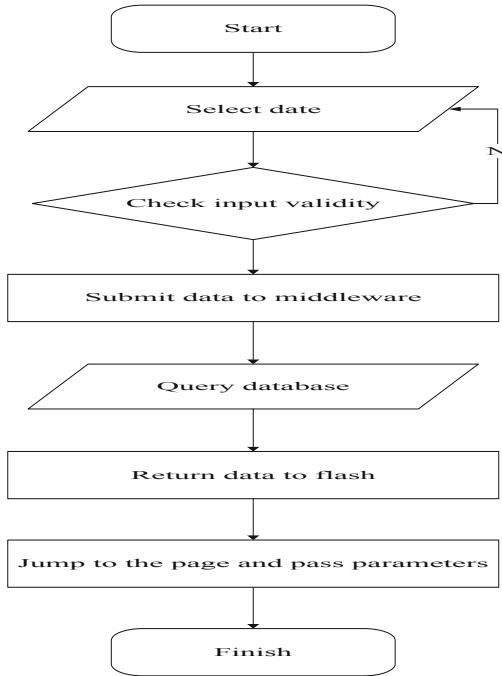


Fig. 1. Date input and data query flow chart

(2) Hotel room selection interface

The interface is designed to create a virtual 3D hotel scene environment in which you can freely rotate and change the perspective. It mainly contains a 360° panoramic view of the surrounding environment of the hotel building, showing a simple floor distribution map of the hotel rooms. Then query the database according to the data sent by the flash, and return the available guest room resources to the user for selection. The user can click on an available guest room and see a brief introduction of the guest room and a dynamic interactive panorama of the guest room. After the user selects a room, enters the relevant reservation information, and confirms with the customer over the phone, the check-in process can be simplified when arriving at the hotel, and the result of the quick check-in can be simplified.

3.4 Message Management Module

In the message module, the administrator checks the user’s comments on hotel services and other information, and deletes unnecessary information by pressing the delete button. In this part, you can view the messages of front-end users and back-end administrators to realize the communication and process between customers and managers.

4 Discussion

In the system test, the system performance test is also a very important link, but compared with the function test, its function is mainly used to verify the system performance index. Performance testing can be run manually or as an automated software test, but the cost of using manual testing is relatively high, so performance testing is usually run using automated software. LoadRunner is a testing tool that recognizes and tests problems by simulating tens of millions of users and conducting or implementing synchronous load characteristic monitoring, and then predicts load behavior and system characteristics. It is widely used in system characteristic detection. This series also uses this kind of software is used to detect system characteristics.

Table 1. System response time and CPU usage

	User module	Room management module	Reservation module	Message management module	CPU usage
50	0.52	0.75	0.61	0.36	11.6%
100	0.81	0.91	0.98	0.57	13.2%
200	1.68	1.79	1.45	0.83	16.7%
300	1.94	2.01	1.82	1.52	18.2%
400	2.35	2.54	2.33	1.96	24.6%
500	2.96	2.85	2.74	2.47	25.7%

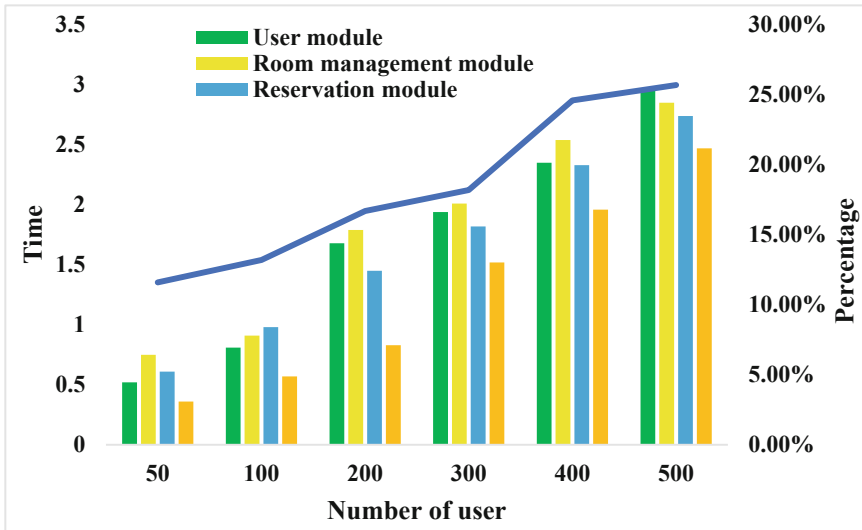


Fig. 2. System response time and CPU usage

It can be seen from Table 1 and Fig. 2 that when the number of concurrent users of the system reaches 500, the response time of each module is 2.96 s, 2.85 s, 2.74 s and 2.47 s respectively, and the response time is within 3 s. The CPU of the system module occupancy rate is also within 30%, which meets the performance requirements of this article.

5 Conclusions

The design and implementation of hotel management information system, its main function is to realize the reservation function of hotel system and implement systematic management of hotel management, which not only saves manpower and material resources, but also makes information management methods more flexible and diverse, and management methods more informative change.

References

1. Davern, D.: The influence of employer branding in talent management in the hotel industry. *Irish Business J.* **12**(1), 3 (2019)
2. Jcn, A., Csy, B., Jkh, C., et al.: Combining non-invasive wearable device and intelligent terminal in healthcare IoT. *Proc. Comput. Sci.* **154**(C), 161–166 (2019)
3. Li, D., Gao, Y., Duan, Z., et al.: Development of intelligent infusion management system. *Zhongguoyiliao qi xiezazhi = Chin. J. Med. Instrument.* **43**(3), 176–178 (2019)
4. Zainudin, F., Ismail, K.: Measuring malaysia school resource centers' standards through iq-pss: online management information system. *IASL Ann. Conf. Proceed.* **2010**(1), 20 (2021)
5. Aldholay, A.H., Isaac, O., Abdullah, Z., et al.: The role of transformational leadership as a mediating variable in DeLone and McLean information system success model: the context of online learning usage in Yemen. *Telematics Inform.* **35**(5), 1421–1437 (2018)
6. Berisha, M., Miftaribasholli, F., Ramadani, N., et al.: Impact of the national population register in improving the health information system of malignant diseases in Kosova. *Acta Informatica Medica*, **26**(1), 62–66 (2018)
7. Ganguly, K., Rai, S.S.: Evaluating the key performance indicators for supply chain information system implementation using IPA model. *Benchmarking Int. J.* **25**(6), 00 (2018)
8. Lavbi, D., Lajovic, I., Krisper, M.: Facilitating information system development with Panoramic view on data. *Comput. Sci. Inf. Syst.* **7**(4), 737–767 (2018)
9. Hudson, D., Zuege, D.J., Kushniruk, A., et al.: Physician satisfaction with a critical care clinical information system using a multimethod evaluation of usability. *Int. J. Med. Inf.* **112**, 131 (2018)
10. Katzan, I.L., Lapin, B.: PROMIS GH (patient-reported outcomes measurement information system global health) scale in stroke: a validation study. *Stroke* **49**(1), 147–154 (2018)
11. Barale, V.: A supporting marine information system for maritime spatial planning: the European Atlas of the Seas. *Ocean Coastal Manage.* **166**, 2–8 (2018)
12. Kasturi, S., Szymonifka, J., Burket, J.C., et al.: Feasibility, validity, and reliability of the 10-item patient reported outcomes measurement information system global health short form in outpatients with systemic Lupus Erythematosus. *J. Rheumatol.* **45**(3), 397–404 (2018)



Fuzzy Sets Qualitative Comparative Analysis (fsQCA) of Helicopter Accident Casualties Based on HFACS Model

Bei Zhou^(✉)

School of Management, Wuhan University of Technology, Wuhan, Hubei, China

Abstract. With the increasing use of civil helicopters, the number and frequency of accidents and air crashes involving helicopters is also on the rise. To explore the factors influencing helicopter accidents and the pathways of low accident casualty generation, based on 27 helicopter accident/accident symptom investigation reports released by the Civil Aviation Administration of China from June 2018 to September 2020, the HFACS model (Human Factors Analysis and Classification System) was used as a theoretical framework and the fuzzy set Qualitative Comparative Analysis (fsQCA) method was applied to explore the complex mechanism and causal asymmetry of multiple concurrent conditions and low accident casualties in the four dimensions of accident occurrence. The fsQCA method was used to investigate the mechanisms and pathways of accident occurrence. Based on the four pathways obtained, it is found that: helicopter safety accident occurrence is a typical multi-conditional complex relationship with multiple concurrent conditions; pilot error or violation is still an important factor in accident generation; in the case of pilot error or violation, the ways to reduce accident casualties are to ensure that the aircraft is fault-free and the terrain and weather are compatible with the flight conditions, and to ensure that unsafe supervision is in place or that organizational management is up to standard.

Keywords: Helicopter · HFACS · Fuzzy Set Qualitative Comparative Analysis (fsQCA) Method

1 Introduction

With the obvious opening of China's control of low-altitude airspace, general aviation has shown leaps and bounds. Helicopters are widely used in the field of aviation because of their special properties such as fast airframe and mobility, which are different from transport aircraft, such as flight training, aerial excursions, agricultural and forestry operations, law enforcement, emergency medical treatment, aerial photography, marine monitoring, search and rescue, etc. According to the "China Helicopter Industry Analysis Report 2020", the number of civil registered helicopters in China is 1,216 by 2019. With the increasing use of civil helicopters, the number and frequency of accidents and air crashes involving helicopters is also on the rise. The analysis of 45 accident/accident symptom investigation reports released by CAAC in the past three years from June 2018

to September 2020 found that 60% of the accidents/accident symptom were helicopter accidents, and helicopters are characterized by many rotating mechanisms, complex aerodynamic environment, and alternating loads on the airframe and components due to their type and operating environment. The factors influencing the occurrence of accidents are complex and have the characteristics of multi-factor coupling.

In recent years, the human factor analysis of accidents has gradually shifted from focusing on human elements to systemic analysis, which should focus not only on human influences but also on the links between human factors and other potential conditions and variables when conducting safety accident analysis. In 2001, Shappell et al. first proposed the HFACS (Human Factors Analysis and Classification System) accident analysis model for 4-level factor analysis of accidents [1]. HFACS provides a systematic classification of accident causative factors for safety accident cases and is a more comprehensive and reliable analysis method in this field, but only a few of the many studies utilizing HFACS have focused on helicopter accidents, for example, Liu et al. used the HFACS model to analyze 717 foreign civil helicopter accidents and summarized the factors of five types of helicopter accidents [2]. Majumdar et al. conducted a statistical analysis of the causes of helicopter accidents in New Zealand 1996–2005 and the UK 1986–2005 [3]. Rao analyzed helicopter maintenance risks through accident data collected [4]. Zhang Juan et al. conducted a statistical analysis of the causes of helicopter accidents in the world from 2014–2016 from several aspects [5]. These studies have derived single significant causes of accidents or ranked primary and secondary causes and have focused on military helicopter accidents and emergency medical aid helicopter accidents.

Since the goal of HFACS is to understand the potential causes of safety incidents and to consider only the upper-level continuous influence on the lower level in the system analysis [6]. Whereas the relationship between unsafe conditions and human unsafe behavior is a one-to-many interaction [7], which leads to a non-linear and complex interaction between human factors and other causal factors. In order to explore the complex paths of action between accident causation, it is indeed necessary to find suitable analytical tools to dissect this nonlinear relationship. In addition, the shortcomings of HFACS in quantitative analysis limit its application, and other tools are needed to assist in causal chain association analysis to explore the interaction mechanisms between human factors and other factors. For example, Hu et al. used a combination of structural equation modeling and HFACS to quantitatively dissect the interrelationships between the various causes [8]. Xia et al. hybrid Bayesian networks to carry out BN-HFACS analysis are able to quantitatively dissect the failure probability and sensitivity factors between safety risk factors and project safety performance [9]. Tang et al. developed an improved HFACS model for personalized safety management, and preliminary results show that the model outperforms traditional safety management by reflecting and analyzing the individualized safety status of workers [10]. While HFACS provides a systematic classification of accident causal factors for safety accident cases, it is necessary to combine other analytical tools in order to explore common trends and recurring problems in the development of helicopter-only accidents, and the use of qualitative comparative analysis, a case study-oriented theoretical pooling research method, can both build a framework for safety accident analysis based on a holistic perspective and break the deficiency

of quantitative content of HFACS in accident causal factor classification by combining qualitative and quantitative means.

2 Model Construction and Data Processing

Qualitative Comparative Analysis (QCA) method combines case studies with quantitative studies and can perform group analysis on different combinations of conditions. Currently, the QCA approach is mainly applied to research in related fields such as public safety, news communication, and corporate innovation and strategic management [11]. An important step in fuzzy sets is the choice of “calibration”, and this paper uses indirect calibration and three-valued fuzzy sets, which require setting three qualitative turning points (1 = fully affiliated 0.67 = biased affiliated 0.33 = biased unaffiliated 0 = unaffiliated).

2.1 Case Selection

With the help of the website of CAAC, a total of 45 aircraft accident/accident sign investigation reports were collected from 2018 to 2020, of which 27 cases were helicopter accident/accident sign investigation reports except for aircraft accident/accident symptom investigation reports of the types of ventilation aircraft, seaplanes, and light sport aircraft, and in the accident type, time and place of occurrence, accident results and other dimensions, showing diversified features, are supported by the accident/accident symptom investigation report, complete information, the accident occurred more clearly.

2.2 Variable Selection and Assignment

QCA emphasizes the identification of conditions relevant to the research question based on theoretical or empirical knowledge. Selecting conditions based on the research framework and theoretical perspectives is the main method used by researchers. In the previous narrative, HFACS theory is widely used in the aviation field and is a more comprehensive and reliable analytical approach. HFACS theory presents four different types of human factors problems, i.e., organizational factors, unsafe supervision, preconditions for unsafe behavior, and unsafe behavior at four levels of failure.

A key factor to consider in constructing a group model is the number of conditions. The number of conditions is determined by considering both the sample size and the parsimony of the model. Since there are theoretically 2^k combinations of k conditions, an excessive number of conditions can easily lead to the number of histories exceeding the number of observed cases and thus the problem of “limited diversity” of cases.

In this paper, based on HFACS theory, the influencing factors of helicopter safety accidents are divided into four specific categories of organizational influence, unsafe supervision, preconditions conditions of behavior and unsafe behavior with a total of five conditional variables (Fig. 1), and the specific content and assignment of each conditional variable are shown in Table 1.

2.3 Analysis of Necessary Conditions

The assigned data were imported into fsQCA software for the necessary condition analysis, and it was usually determined that the necessary conditions needed to reach a consistency score of 0.9 and have sufficient coverage. The results of the necessary condition analysis are shown in Table 2, and the consistency values of all the conditions were less than 0.9.

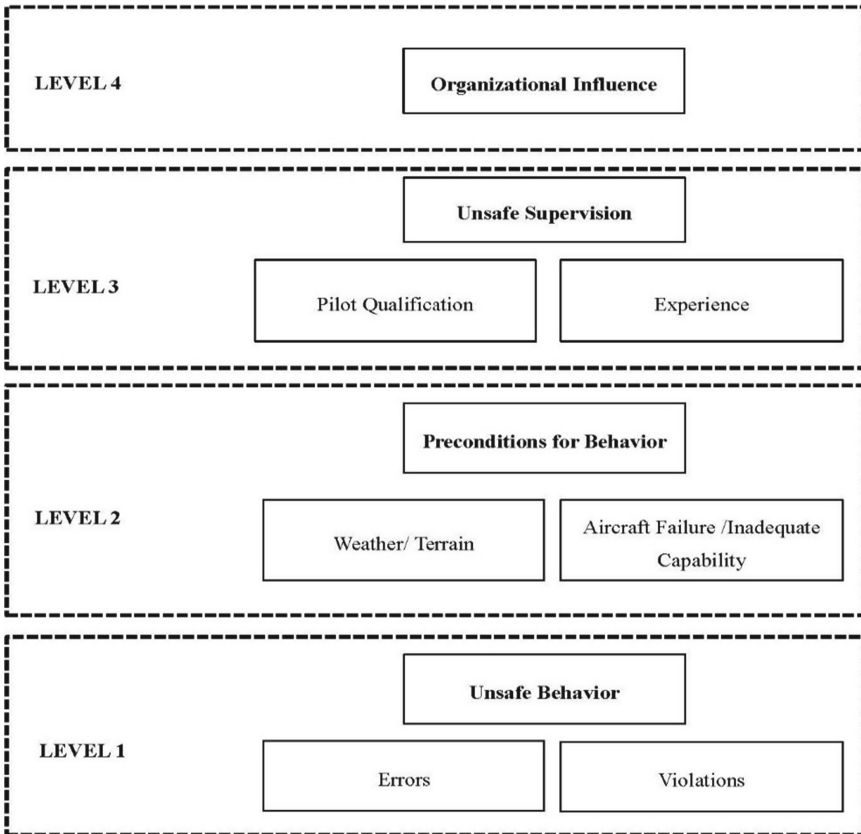


Fig. 1. HFACS model

2.4 Truth Table and Conditional Combinatorial Conformation Analysis

In this paper, accident investigation reports were used as the source of cases, and the results were negative, i.e., there were injuries or deaths, so the combination of conditions that could produce “low injuries or deaths” was studied here. The truth table was constructed with a minimum number of cases of 1 and a threshold of 0.8, and three solutions were obtained from the standard analysis (Tables 3, 4, and 5): complex, parsimonious,

and intermediate solutions, and the results of the statistical analysis showed that the consistency value of the overall solution exceeded 0.80 and the consistency score of the solution exceeded 0.80, indicating that the explanation of the outcome variables was good.

Among the three solutions, the intermediate solution is representative, and the intermediate solution is selected for analysis, supplemented by the parsimonious solution to identify the core and edge conditions, and the low injury/death formation paths in the sample cases are shown in Table 6.

Table 1. Interpretation and assignment of variables

Variable name				Assigning values	
Conditional variables	Unsafe behavior	Errors/violations	E/V	Errors and violations	1
				Errors or violations	0.67
				Errors/violations not excluded	0.33
				No errors or violations	0
	Preconditions for unsafe behavior	Weather/terrain	W/T	Both have an influence	1
				One of them have an influence	0.67
				Not excluded	0.33
				Neither the environment nor the terrain has any influence	0
		Aircraft failure/inadequate capability	AF	Aircraft failure	1
				Aircraft is trouble-free	0
	Unsafe supervision	Pilot qualification/experience	Q/E	Qualifications/experience qualified	1
				Qualifications/experience not qualified	0
Organizational influence	Management	MANAGE	Organizational management up to standard	1	
			Inadequate organizational management	0	
Outcome variable	Accidents	Dead/injured	D/I	More than 2 deaths	1
				Death of 2 people	0.67
				Death of 1 people	0.33
				Injuries only	0

Table 2. Analysis of necessary conditions

Outcome variable: ~D/I		
Variable	Consistency	Coverage
E/V	0.605247	0.828205
~E/V	0.539038	0.564052
W/T	0.833229	0.580000
~W/T	0.166771	0.667500
AF	0.103073	0.443333
~AF	0.896927	0.611667
Q/E	0.625234	0.651693
~Q/E	0.519051	0.713917
MANAGE	0.270456	0.618571
~MANAGE	0.729544	0.584000

Table 3. Complex solution

Model: ~ D/I = f(E/V, Q/E, AF, W/T, MANAGE) frequency cutoff: 1. consistency cutoff: 0.800399

	Raw coverage	Unique coverage	Consistency
E/V*~Q/E*~AF*~W/T	0.125547	0.125547	1
E/V*Q/E*W/T*~MANAGE	0.312929	0.251093	0.833611
E/V*Q/E*~AF*MANAGE	0.124922	0.124922	0.854701
~E/V*Q/E*~AF*~W/T*~MANAGE	0.186134	0.124297	0.818681

Solution coverage: 0.687695
 Solution consistency: 0.868297

Table 4. Parsimonious solution

Model: ~ D/I = f(E/V, Q/E, AF, W/T, MANAGE) frequency cutoff: 1. consistency cutoff: 0.800399

	Raw coverage	Unique coverage	Consistency
E/V*W/T	0.37539	0.251093	0.857347
~Q/E*~W/T	0.146159	0.0206121	0.78
E/V*~Q/E	0.125547	0	1
E/V*MANAGE	0.166771	0.062461	0.887043
~E/V*~AF*~W/T	0.309806	0.144909	0.833613

Solution coverage: 0.728919
 Solution consistency: 0.833571

Table 5. Intermediate solution

Model: $\sim D/I = f(E/V, Q/E, AF, W/T, MANAGE)$.consistency cutoff: 0.800399			
	Raw coverage	Unique coverage	Consistency
$E/V * \sim Q/E * \sim AF * \sim W/T$	0.125547	0.125547	1
$E/V * Q/E * W/T * \sim MANAGE$	0.312929	0.251093	0.833611
$E/V * Q/E * \sim AF * MANAGE$	0.124922	0.124922	0.854701
$\sim E/V * Q/E * \sim AF * \sim W/T * \sim MANAGE$	0.186134	0.124297	0.818681
Solution coverage: 0.687695			
Solution consistency: 0.868297			

Note: The core condition is the variable that appears in both the parsimonious solution and the intermediate solution, and the variable that appears only in the intermediate solution but not in the parsimonious solution is the edge condition.

As can be seen from the table, a total of four paths are generated. Path 1 can be expressed as $E/V \times \sim Q/E \times \sim AF \times \sim W/T$; Path 2 can be expressed as $E/V \times Q/E \times W/T \times \sim MANAGE$; Path 3 can be expressed as $E/V \times Q/E \times \sim AF \times MANAGE$; Path 4 is expressed as $\sim E/V \times Q/E \times \sim AF \times \sim W/T$.

After adjusting the threshold values to 0.78 and 0.7, respectively, the path did not change, indicating that the result is stable.

Table 6. Low D/I formation pathways (N = 27)

Variable	H1	H2	H3	H4
E/V	●	●	●	⊗
Q/E	⊗	●	●	●
AF	⊗		⊗	⊗
W/T	⊗	⊗		⊗
MANAGE		⊗	●	
Consistency	1	0.833611	0.854701	0.818681
Coverage	0.125547	0.312929	0.124922	0.186134
Solution Coverage: 0.687695				
Solution Consistency: 0.868297				

3 Discussion and Suggestion

From the above analysis, it is clear that helicopter accidents are the result of multifactorial coupling, and when organizational management, environment, and unsafe supervision are in place, the severity of accidents is reduced even when pilots have errors and violations.

In path 1, when the pilot has errors or violations and unsafe supervision is insufficient, ensuring that the aircraft is fault-free and the terrain and weather meet the flight conditions can effectively reduce the severity of the accident, for example, case 9 occurred

on April 25, 2019 is a flight training accident, because the instructor let go of too much and the trainee with only 2 h of flight experience, when the helicopter sled touched the ground, there was a subconscious jerking the cycle to the right. However, there was no other mechanical failure of the aircraft, and the weather conditions were good and the place was flat, so the accident did not cause casualties; in path 2, when there is pilot error or violation, insufficient organization and management, and the terrain or weather has influence on the flight, ensuring unsafe supervision in place can also effectively reduce the severity of the accident, such as July 30, 2018, the helicopter yawed to the right but no one on board was aware of it, and the company did not fully consider the captain's ability to handle special situations, and there were deficiencies in organization and management, but the controller was alerted and the copilot gave a reminder, and the accident only caused four injuries and no one died; in path 3, when The accident severity is lower when the pilot has errors or violations, unsafe supervision is in place and the organization and management are up to standard, and the aircraft is not faulty. Path H4 indicates that the accident severity is lower when there is no pilot error or violation, unsafe supervision is in place, the aircraft is trouble-free and the terrain and weather meet the flight conditions.

4 Conclusion

In this paper, we analyze 27 helicopter accident cases based on HFACS theory and fsQCA method, and get a total of four paths of low accident casualties, and analyze the core and edge conditions of these four paths, and put forward relevant analysis and suggestions for the paths. This study focuses on the main paths of helicopter accidents with low accident casualties, and proposes corresponding blocking measures according to the characteristics of different types of accidents, but has not yet verified the intervention effects of different measures.

References

1. Shappell, S.A., Wiegmann, D.A.: The human factors analysis and classification system—HFACS. *Hum. Factor. Aerosp. Safety* **1**(1), 59–86 (2001)
2. Liu, G., Xiao, H., Yu, L.-H.: Research on human cause and fatality of civil helicopter accidents. *J. Safety Sci. Technol.* **12**(9), 133–137 (2016)
3. Majumdar, A., Mak, K., Lettington, C., Nalder, P.: A causal factors analysis of helicopter accidents in New Zealand 1996–2005 and the United Kingdom 1986–2005. *Aeronaut. J.* **113**(1148), 647–660 (2009)
4. Rao, A., Fala, N., Marais, K.: Analysis of helicopter maintenance risk from accident data. In: *Proceedings of American Institute of Aeronautics and Astronautics Science and Technology Conference, San Diego, 4–8 January 2016*. American Institute of Aeronautics and Astronautics, Reston (2016)
5. Zhang, J., Zhan, Y., Wang, Y.: Statistics and analysis of world helicopter accidents in 2014–2016. *Helicopt. Technol.* **3**, 68–72 (2017)
6. Sulaiman, M., Samiullah, I., Hamdi, A., et al.: An improved whale optimization algorithm for solving multi-objective design optimization problem of PFHE. *J. Intell. Fuzzy Syst.* **37**(3), 3815–3828 (2019)

7. Nobahari, M., Seyedpoor, S.M.: An efficient method for structural damage localization based on the concepts of flexibility matrix and strain energy of a structure. *Struct. Eng. Mech.* **46**(2), 231–244 (2013)
8. Hu, S., Li, Z., Xi, Y., et al.: Path analysis of causal factors influencing marine traffic accident via structural equation numerical modeling. *J. Marine Sci. Eng.* **7**(4), 96 (2019)
9. Xia, N., Zou, P.X.W., Liu, X., Wang, X., Zhu, R.: A hybrid BN_HFACS model for predicting safety performance in construction projects. *Safety Sci.* **101**, 332–3434 (2018)
10. Tang, N., Hu, H., Xu, F., et al.: A personalized Human Factors Analysis and Classification System (HFACS) for construction safety management based on context-aware technology. *Enterp. Inf. Syst.* **16**(1), 141–166 (2021)
11. Pot, W.D., Dewulf, A., Biesbroek, G.R., et al.: What makes decisions about urban water infrastructure forward looking? a fuzzy-set qualitative comparative analysis of investment decisions in 40 Dutch municipalities. *Land Use Policy* **82**, 781–795 (2019)



Animation Game Database Model Based on Virtual Reality Technology

Huihuang Wu^(✉)

Xiamen Software Vocational and Technical College, Xiamen 361024, Fujian, China

Abstract. As an emerging cultural and economic phenomenon, animation industry is closely related to human life, and its development process is also affected by factors such as environment and talents. With the continuous progress of Internet technology, the popularization of computer network and mobile communication, virtual reality technology provides us with a new idea. It can realize its function and value goal by establishing a model, and can combine the traditional design with modern advanced software for development, so as to make the whole database more intuitive, humanized and intelligent. Therefore, this paper designs the animation game database model based on virtual reality technology. Firstly, this paper introduces the concept and interactive characteristics of animation game database, then introduces the application of virtual reality technology, and designs the animation game database model. Finally, the running performance of the model on different types of computers is tested. The test results show that the frame rate of the game running picture of the game laptop has reached 20 frames per second, basically reaching the target frame rate. After the evaluation and feedback of different players, there is no picture pause and discomfort. The game content runs smoothly and has a good experience.

Keywords: Virtual reality technology · Animation games · Data base · Model design

1 Introduction

The development of animation industry is based on animation and applies virtual reality technology to three-dimensional space. In the period of rapid economic growth and high Internet penetration in China, people have higher requirements for material living standards, while the traditional two-dimensional plane information model has been unable to meet people's needs for vision, hearing and touch [1, 2]. This makes 3D image library one of the most widely used forms in the digital media market in the new era, and its development prospect is immeasurable [3, 4].

Many scholars at home and abroad have studied virtual reality technology. In recent years, with the rapid development of computer technology, people gradually realize the importance of computer. Virtual reality technology is used in the database to build a complete, efficient and operable software system. On the basis of studying the current situation in China, we can find that foreign countries have formed a mature theoretical system on the database model of virtual reality technology. However, there are still

deficiencies for the animation industry. China is currently in the development stage of stand-alone Internet. Some scholars believe that China should strengthen the construction of online game database system [5, 6]. Some scholars pointed out that we should strengthen the research on game development and solve the application of network virtual reality technology in animation industry in combination with the actual situation [7, 8]. Although these scholars have discussed it to a certain extent, there are still deficiencies. Therefore, based on virtual reality technology, this paper further studies the animation game database model.

This paper mainly studies the animation database model based on virtual reality technology. Firstly, it defines the related concepts, and expounds the differences and their unique characteristics between 3D games and traditional 2D works. Secondly, it analyzes and compares different types of online games and web page design cases at home and abroad, database system and problems encountered in the process of web page production. At the same time, it puts forward countermeasures and suggestions combined with the domestic current situation. Finally, it summarizes the conclusions and looks forward to the future development prospects.

2 Discussion on Animation Game Database Model Based on Virtual Reality Technology

2.1 Animation Game Database

2.1.1 Concept

The construction of animation database is mainly the overall design and analysis of an animation system. After the overall scheme is determined, the whole system is realized through certain software. The design of animation database model is based on the animation production process. After analyzing the whole game process, it determines what functions the system wants to achieve according to the relevant data information. First of all, we should grasp every detail as a whole. Each module has its own unique characteristics and functions. Secondly, it makes a detailed description of what to do and how to do in each stage. Finally, it designs the database table structure diagram and specific operation methods. In this process, it is mainly analyzed and studied in combination with the game production process [9, 10].

2.1.2 Interaction Characteristics

In the system, it mainly communicates between users and viewers, and realizes the mutual communication, interaction and information transmission between viewers and administrators in the virtual environment through role movement, scene setting and virtual reality technology. In the database, the whole process of the game is completed by the combination of animation and virtual reality technology, and the most important is the role model, scene and the relationship between scenes. The data in the database refers to all the original materials required for the game development process, including user demand analysis and design documents. During the game system, its specific functions can be determined according to different characters, scenes and other information [11, 12].

2.2 Virtual Reality Technology

2.2.1 Principle

Progress is the inevitable result of the development of the times. In the information environment, people also put forward higher requirements for the animation industry. Virtual reality technology provides a good platform for it. The research of virtual reality technology started in the 1980s. It is a new discipline developed in the field of computer. With the continuous progress of science and technology such as Internet and mobile communication, people are more and more dependent on the network environment. Virtual reality technology is based on computer and carries out digital processing in three-dimensional space. Use the computer to simulate the environment of the object, place the object in the three-dimensional space, and simulate an entity through the computer to make it have a certain degree of dynamic response. The corresponding animation model base system is established according to the character model. Then, after the character modeling is completed, it is applied to the actual operation to realize the functional modules such as rendering effect and motion trajectory. Finally, computer technology is used to create solid objects in three-dimensional space and construct virtual space, so as to simulate the real world environment and make it more realistic to reflect the real world.

2.2.2 Virtual Scene Generation

According to the different methods of making virtual environment, the virtual reality technology through software can be divided into graphics based virtual reality technology and model-based virtual reality technology, which have their own advantages and disadvantages. The model-based method also includes establishing a virtual environment with geometric units, which can not only depict geometric elements through computer graphics technology, but also through its own modeling tools, such as AutoCAD, 3D studio, etc. It enhances the authenticity of virtual reality scene and can be presented in real time. Once the virtual reality model is established, the interactive virtual environment can be realized by adding operations such as moving, flipping and angle transformation to the event response. Through the model, it can independently complete the construction and rendering according to the attributes of physical objects, and has a large degree of freedom. However, due to the characteristics of model-based approach, it also faces some problems. Virtual reality technology using image modeling refers to extracting image sequences by using collected discrete images or videos, composing image sequences into coherent panoramic graphics, and then organizing multiple panoramic images in the virtual panoramic room through an appropriate spatial model. The user retrieves the panoramic image from the storage medium, and can view the virtual panoramic space and move along a fixed path.

2.2.3 Image Preprocessing

The samples used in modeling need to be obtained by real-life cameras. When the lens or camera device of the camera system is not facing the scene to be photographed, the photographed scene picture will also change accordingly. Of course, this is inevitable.

Then, how to reduce the geometric distortion of the digital image generated by the camera is the main problem to deal with. For any geometric distortion, it can be explained by the relationship between the original image position and the distorted image position. Let the coordinates of the original drawing be (x, y) and the coordinates of the distorted drawing be (x', y') , so the relationship between the two positions can be explained by the following equation:

$$\begin{aligned}x' &= h_1(x, y) \\y' &= h_2(x, y)\end{aligned}\quad (1)$$

If $G(x, y)$ is used to represent the gray level of the original image at point (x, y) and $f(x', y')$ is used to represent the gray level of the distorted image at point (x', y') , there should be:

$$g(x, y) = f(x', y') \quad (2)$$

Thus, the problem of recovering the restored image by eliminating geometric distortion is reduced to the problem of obtaining $g(x, y)$ from the relationship H1 and H2 between the distorted image $f(x', y')$ and the two coordinates.

3 Experiment

3.1 Establishment of Game Database Development Environment

The development of animation database management system is a systematic project. On this platform, we need to make a reasonable planning for its construction environment, taking into account the interaction between game screen and interface, followed by technology. The establishment of animation database model is mainly designed for players of different types and genders. According to the game scene and user needs, a virtual reality technology system is established in the database. First, we need to analyze the whole network environment. By understanding the characteristics of all information on the network and its corresponding relationship. Second, determine whether there is a conflict between the required data source and storage resources. Finally, select appropriate parameters to build the model library, design a virtual reality technology model library that can meet the user's functional needs, convenient operation and other needs, and has good scalability, and finally realize the construction and maintenance of the game database model, so as to provide some technical support for subsequent development.

3.2 Specific Structure of Animation Database

Figure 1 is the structure diagram of animation game database based on virtual space technology. The construction of game database model is mainly aimed at animation works. According to different types and themes, the corresponding database structure is established. The database table is established in virtual reality technology. The form includes: role ID, registration button, etc. Players click the mouse to determine the

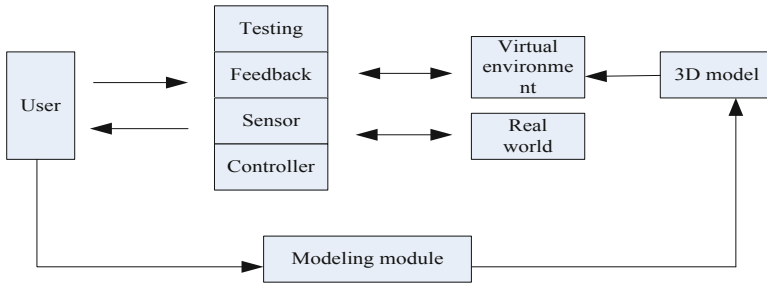


Fig. 1. Animation and game database structure

information of others within their search range. Then connect it with the records in the database table, and finally return to the home page for the next operation, so as to realize the circular process of the final target function module and complete the game library, that is, connect the animation data management and user login interface.

3.3 System Test Process

The system performance test mainly includes the following aspects: (1) data processing ability. After inputting data in the database, the corresponding response can be made quickly. For different types of users, they use different models. For example, users need to log in to manage and change their passwords. They must first register and save the information, and then send it to the administrator or manager to check whether there is a problem. If there is no exception or the system has been activated, they cannot log in to the database again to process and analyze the data. (2) Interface design. After registering as a member, users can enter the website page to browse relevant information and download animation works. When users are using, they cannot provide services to others. When you cannot log in to the website interface, you need to set a special function above it or exit the corresponding permission to allow you to join.

After the performance test of the system is completed, a series of operations need to be carried out, including data acquisition, saving and transmission, storage and so on. Because the database itself is unstable, we use MySQL as the background server to manage and control all virtual objects in the animation library. At the same time, we can also view the animation playback progress on each page and the display of special effects of each part through the network query function. In addition, links about animation content can be added to the web page to facilitate users to understand relevant information and modify materials.

4 Discussion

4.1 Running Performance Test of Animation Game Database

Table 1 is the performance test data table of animation game database based on virtual reality technology.

Table 1. Run the performance test data

Computer model	Game operation time (ms)	Game delency time (ms)	Frame drop rate (FPS)	Memory occupancy (%)
Game laptop	56	14	20	1
Desktop computer	123	247	78	3
Work laptop	134	253	50	2

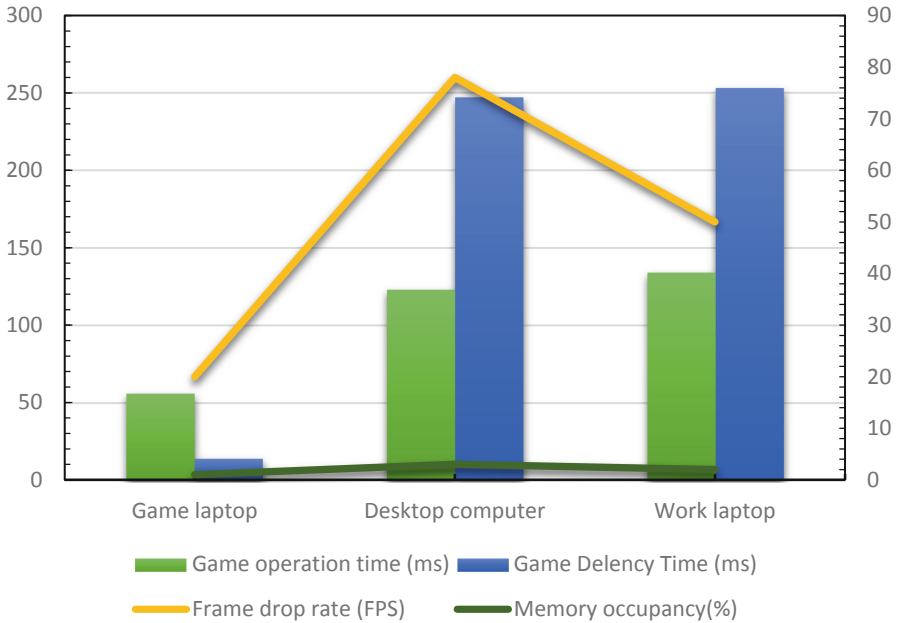


Fig. 2. Database run test comparison

The system performance test is mainly aimed at the running environment of the database, and the simulation experiments are carried out under different application scenarios. Through the analysis of data and parameter changes to verify whether the designed software can achieve the expected function. It can be seen from Fig. 2 that the test results of different hardware configurations. The game running picture frame rate of the game laptop used in the game development in this paper has reached 20 frames per second, basically reaching the target frame rate. After the evaluation and feedback of different players, there is no picture pause and discomfort. The game content runs smoothly and has a good experience.

5 Conclusion

This paper mainly focuses on the animation database model, makes full preparation and analysis of the preliminary work, and completes a complete animation database system combined with relevant cases, references and existing materials. In this paper, a series of application experiments are carried out for virtual reality technology. Through consulting materials and online search, some research results and latest trends in the field of 3D scene development and maintenance at home and abroad in recent ten years, and through a series of operation performance tests of game databases, the experimental results show that the game content runs smoothly and has a good experience.

References

1. Junjie, Y.: Research on the application of 3D animation and virtual reality technology. *Inf. Comput.* **2**, 190–191 (2017)
2. Li, Y.M., Joo, J.W.: The change of animation narrative structure caused by virtual reality technology. *J. Dig. Cont. Soc.* **20**(3), 459–467 (2019)
3. Wei, C., Hongke, L.: Research on scene simulation based on the virtual reality technology - Taking Middle School Geography Teaching for example. *Mod. Educ. Technol.* **27**(11), 19–25 (2017)
4. Gan, B., Zhang, C., Chen, Y., Chen, Y.-C.: Research on role modeling and behavior control of virtual reality animation interactive system in Internet of Things. *J. Real-Time Image Proc.* **18**(4), 1069–1083 (2020)
5. Zhou, R.: Research on information management based on image recognition and virtual reality. *IEEE Access*, **99**, 1 (2020)
6. Zhang, X., Yin, Y.: Design of training platform for manned submersible vehicle based. In: *Proceedings of the 31st International Conference on Computer Animation and Social Agents (CASA 2018)*, Beijing, China, 21 May 2018 to 23 May 2018, pp. 90–94. ACM Press (2018)
7. Liang, M.C., Yin, Y., Liu, H., et al.: Simulation of ROV retractable operation based on virtual reality technology. *Dalian Haishi Daxue Xuebao/J. Dalian Maritime Univ.* **44**(4), 49–54 (2018)
8. Li, S., Hou, Y.: Study on the architecture design and interior decoration based on VR technology and computer simulation platform. *Paper Asia* **35**(2), 54–57 (2019)
9. Jeon, G.: Study on the visual characteristics and subjectivity in the live action based virtual reality. *Catoon Anim. Stud.* 117–139 (2017)
10. Farahani, J., Soltani, P., Roberts, R.E.: Expertise differences in a 2D animation simulation decision-making task: the influence of presentation speed on performance. *Prog. Brain Res.* **253**, 87–100 (2020)
11. Suh, D.-H.: Analysis of serious game elements of the contents for smart device based first-aid education. *Catoon Anim. Stud.* 273–294 (2017)
12. Li, W., Cao, M., Wang, Y., Tang, C., Lin, F.: Mining pool game model and nash equilibrium analysis for PoW-based blockchain networks. *IEEE Access* **8**, 101049–101060 (2020)



The Application of Drones in Urban Fire Monitoring

Gang Wang^(✉), Yang Cao, and Jun Zhou

Taizhou Institute of Science and Technology, Nanjing University of Science and Technology,
Taizhou, Jiangsu, China
wanggang@njjust.edu.cn

Abstract. As a multifunctional high-speed aircraft, small drones carry a variety of auxiliary equipment to perform a variety of tasks, and have great practical value in both military and civilian aspects. The traditional fire detection method has the problem of blind spot detection, high cost and poor real-time performance. Compared with traditional fire detection methods, the UAV platform fire detection system has the advantages of more flexibility, easier control, and larger space, which is very important for fire prevention and control. This paper studies the application of drones in urban fire monitoring. It understands the relevant theories of urban fires on the basis of literature data, and then designs the drone fire detection platform, which is designed to improve the performance of the detection platform, optimize the positioning algorithm of the platform, and use experiments to test the optimized algorithm. The detection result shows that the LADV-Hop algorithm performs positioning refinement for nodes that are prone to accumulate errors when solving their own positions, and selects local beacons Nodes perform calculations to avoid the impact of accumulated errors. Therefore, in the case of the same number of nodes as a whole, the positioning error of the SLADV-Hop algorithm is lower than that of the other three algorithms.

Keywords: Unmanned aerial vehicle · Urban fire · Fire monitoring · Positioning algorithm

1 Introduction

Fires are easy to happen, not only very secretive, but also very destructive [1, 2]. According to statistics, among many disasters, the direct losses caused by fires are about three times that of floods, second only to earthquakes and droughts. Therefore, in fact, fires are the primary threat to human life and the most serious threat [3, 4]. For a long time, people have relentlessly learned many methods of extinguishing fire, preventing and eliminating hidden dangers [5, 6]. In recent years, with the continuous progress and development of science and technology, people have used the latest high-tech methods to study the mechanisms and laws of fire occurrence, development and prevention, and developed new fire protection concepts and technologies to make fire extinguishing technology more scientific and systematic [7, 8].

Regarding fire monitoring research, some researchers said that traditional forest fire monitoring methods include observation decks and human observers. With the new development of technology, in recent decades, forest fire monitoring is mainly based on ground fire monitoring, man-machine monitoring and satellite monitoring. However, each method has some different problems. For example, the ground surveillance system is usually fixed at a specific location, and the surveillance range is very limited. Man-machine monitoring is usually very expensive. At the same time, the dangerous flight environment, bad weather and the physical conditions of the earth threaten the lives of pilots. The cost of installing satellite systems is high, installation and technical update requirements are strict, and sometimes even closer fire monitoring is required, and detailed data needs to be collected. Satellite monitoring cannot meet this collection requirement [9]. Some researchers have also proposed a method for night video fire detection based on four-seed algorithm and data fusion method, and make decisions by combining sub-algorithms. According to the characteristics of fire, they developed an entropy-based method, which is an online adaptive fire detection decision-making framework, and used existing fire videos to verify the effectiveness of the algorithm [10]. Although various methods of monitoring forest fires are mature, few experiments use drones to monitor forest fires. Because in actual use, there are still many challenges, such as how to choose the right platform and UAV sensors, autonomous navigation and control algorithms, and how to combine UAVs with other remote sensing technologies.

This article explores the application of drones in urban fire monitoring. It analyzes the characteristics of urban fires and the application of drones on the basis of literature, and designs the drone fire monitoring platform, and then the design platform is tested, and the relevant theories are derived from the test results.

2 Urban Fire Research

2.1 City Fire Characteristics

Fires usually occur in potentially high-risk areas in cities. When the solid fuel ignites spontaneously with a heat source of appropriate intensity or internal combustion, the fuel burns first [11]. Combustion at this stage is the slow burning of a substance without visible light, usually manifested as gaseous fuel, smoke, and temperature rise. As the temperature increases, the pyrolysis rate increases and the combustible gas concentration increases. When the temperature or concentration reaches a certain value, smoldering becomes combustible and burns, creating an open flame and igniting the surrounding combustibles, causing a fire.

Since there is no flame in the fuel combustion stage, it cannot be effectively identified in the normal color image, but the flame in this stage has the characteristic of increasing temperature all the time, so it is used to identify the infrared image in the temperature gradient method [12].

When the flame is produced, the fuel and the surrounding air continue to produce the flame. At this stage, the flame glows and generates heat during the combustion process, and at the same time has certain flame dynamic and static characteristics. The static characteristics of the flame in the infrared field of view are characterized by the internal temperature of the flame center up to 1000K, and the brightness of the flame area is

significantly higher than the pixels in other areas. In the video image, the isotherm forms a closed area in the flame area, and the isotherm is represented by contour lines of different temperature ranges in the flame.

The dynamic characteristics of the flame in the infrared field of view are characterized by flame flicker due to turbulence in the airflow. Therefore, the developing flame flickers at a frequency of 10 Hz to 25 Hz, and the flame profile and height continue to change. By combining the above-mentioned dynamic and static properties, most infrared sources can be used at this stage to realize flame recognition. Therefore, the continuous rapid rise in temperature and contour instability are important criteria for determining the location of the fire and the high-risk fire. What they have in common is that in the infrared video image, the color, brightness, and contour of two adjacent frames are different between frames, so they can be distinguished by designing the temperature gradient threshold method and the inter-frame dynamic method.

Compared with other infrared sources, high-risk fire spots are characterized by a rapid, continuous and steady rise in temperature. Therefore, in the infrared video image, the signal amplitude change trend in this area is more obvious than in other areas.

2.2 Application of Drones in Fire Disaster Relief

UAVs are equipped with a variety of sensors to perform emergency rescue missions that incorporate the value of UAV applications. Emergency drone rescue is mainly used for forest fires, urban fire fighting and emergency rescue after disasters. From the standpoint of preventing wildfires, it is equipped with a camera assembly consisting of two cameras. One tilts forward in the direction of flight to find the target, and the other tilts down or down, placing it vertically on the left side, slightly tilted. It is used to confirm and detect suspicious target points in the flying direction of the hovering aircraft, or to monitor the situation in the disaster area where a fire occurs. In terms of urban fire protection, drones can fly at specific heights and positions, so that they can observe the entire fire area at a majestic height in the air, and complete the data collection of fire and rescue scenes, fire overview, area, speed and other data that quickly identify the spread of data in real time, go to the ground control center, analyze the fire source and intensity through infrared images, and conduct on-site life detection. The ground can dispatch rescues scientifically based on this information and formulate rescue plans. The fire department can quickly and efficiently organize and deploy fire brigades according to the rescue plan, thereby improving the efficiency of fire fighting activities and preventing the sacrifice of disaster relief personnel. In terms of post-disaster rescue, multiple drones can cooperate with multiple equipment to complete disaster intelligence collection, 401 communication relay, and post-disaster humanitarian search and rescue.

3 Design of Drone Fire Detection Platform

3.1 Platform Design Framework

The framework of the platform can be divided into three parts: wireless sensor network part, early warning system part, and UAV cloud system part. The framework is shown in Fig. 1.

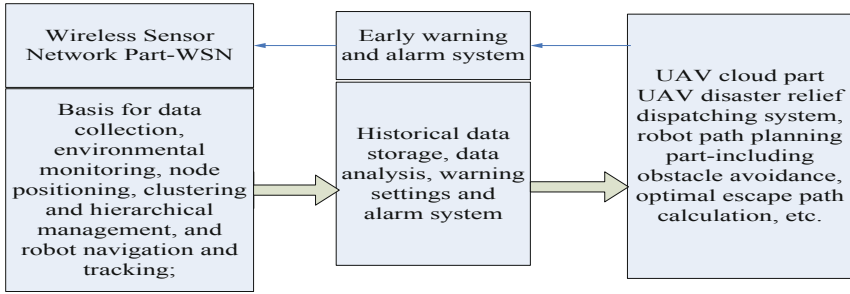


Fig. 1. The frame structure of the system

The relationship between the various modules is roughly described as follows: 1) On the one hand, the wireless sensor network component creates a scalable self-organizing network, collects fusion data and sends it to a higher user level. At the same time, on the other hand, after a disaster, drones participate in rescue operations and provide positioning and navigation functions. 2) In the early warning system part, inspectors can manually start the monitoring network and upload it to this part for data collection. On the other hand, when a fire occurs, the drone part of the cloud system sends an alarm command. 3) As part of the UAV cloud system, after a disaster occurs, a batch of randomly distributed nodes will be sent to the sensor network to form a new self-organizing network. The batch will be as follows: Rescue robots perform fire rescue activities, but when the disaster relief is completed, it will send an alarm cancellation command to the alarm system.

3.2 Image Acquisition Equipment

Taking into account the natural phenomenon of early fires and the cost and stability of the system, the collector of this system uses infrared image sensors and uses complementary semiconductor metal oxide image sensors to collect thermal images. Thermal images use the principle of infrared images. On the other hand, optimal visibility of color images to identify hot spots in the real environment.

3.3 Image Processing

The frame rate of the infrared thermal imaging camera used in this article is 8.3 Hz, which is used to collect infrared images, and the processor used in this article is the DM3730 processor. Experiments show that only the ARM core of DM3730 is used for infrared image processing, and the frame rate can only reach 3 or 4 Hz, which is far from meeting the requirements of real-time processing and monitoring. Therefore, DM3730 with DSP core is required to speed up the processing speed and provide real-time processing and monitoring.

3.4 Network Transmission Equipment

The network transmission equipment used in this article is a high-power wireless remote router, which requires two in total. One is installed on the aircraft as a transmitter, and the other is installed on the ground as a receiver. The integrated system is installed on a small unmanned helicopter. The integrated system is connected to the transmitting router through a wired network, the ground computer is connected to the receiving terminal through a wired connection network, and the integrated system and the ground station can communicate with each other through a local area network. It has two functions. The first is that the computer on the ground can remotely connect to the embedded system through SSH to check the integrated system in the air. The second is to be able to send some monitoring data to the ground station in real time, such as the position coordinates of the firepower target and some parameters of the PTZ.

3.5 Positioning Algorithm Optimization

The DV-Hop algorithm calculates the average bounce distance of each bounce based on the information between the beacon nodes and executes a new transmission circle. The average hop value of the unknown node uses the minimum hop value obtained in the first step as a reference. The hop value is selected as the hop value of the system and multiplied by the hop number of the corresponding beacon node to obtain the distance between the targets. In this way, only the average distance per hop of each anchor node is obtained, so the estimated distance is far from the actual distance.

In response to this type of problem, this article summarizes the average bounce distance and average value of all beacon nodes in the network (such as Eq. (1)), or assigns weights based on the bounce size and average value. For example, formula (2)-the difference between the actual distance of the beacon node and the estimated distance corrects the average bounce distance, and the corrected value is regarded as the bounce distance value of the placed node. These methods can eliminate the average bounce distance error of a single anchor node, but increase the amount of calculation caused by the integration of the entire network beacon node, and continue to accumulate the switching error of the long-distance beacon node between anchor nodes. When estimating the distance, it will speed up power consumption and increase the overall error.

$$HopsIze = i = \sum_{i=1}^K HopSize_i \tag{1}$$

$$HopSize^o = HopSize_i - \sum_{i \neq j} d_{ij}^* - d_{ij} \tag{2}$$

Among them, 1 is the average hop distance of unknown nodes; k beacon nodes; 2 is the average hop distance per beacon node; 3 and 4 are the beacon nodes and the actual distance and estimated distance respectively.

4 Algorithm Test

In order to evaluate the performance of the improved DV-Hop positioning algorithm in this article on wireless sensor networks, in the NS-2 network environment, the traditional DV-Hop algorithm, Selective3-AnchorDV-Hop algorithm, and hop correction. The DV-Hop algorithm proposed in this article will be tested.

The main parameters of the experimental network model are: a 2D plane $200\text{ m} \times 200\text{ m}$ is set up for the development of all sensor nodes in the UAV, the communication radius r of all sensor nodes in the network is the same as 35 m , and the nodes self-organize through the corresponding protocol stack and are in the same twenty experimental simulations were carried out in the environment. The above algorithms were compared and analyzed from the aspects of network node density change, node communication radius r , and the proportion of beacon nodes in the network. Table 1 shows the relationship between the total number of wireless sensor nodes and the positioning error.

Table 1. The relationship between the total number of wireless sensor nodes and the positioning error

	DV-Hop	S3-ADV-Hop	HDCDV-Hop	SLADV-Hop
110	79%	78%	70%	69%
120	70%	60%	56%	45%
130	65%	55%	48%	38%
140	58%	51%	42%	34%
150	55%	48%	35%	28%
160	53%	47%	33%	27%
170	52%	46%	32%	26%
180	51%	45%	31%	25%
190	50%	44%	30%	24%

It can be seen from Fig. 2 that the traditional DV-Hop algorithm and the Selective3-Anchor DV-Hop algorithm do not change the number of bounces and jump distances between nodes, so the placement errors of the two algorithms are basically the same at first. Because the node itself has remaining errors, the DV-Hop algorithm for hop rebound distance correction is passed to the node at the next solution position. Correcting the number of hops and hop distances on a node cannot eliminate the accumulated error when analyzing the position of the node. The SLADV-Hop algorithm selects the local beacon node for calculation to improve the position of the node that is prone to accumulative error when solving the position, and avoid the influence of accumulative error. Therefore, for the same number of nodes as a whole, the placement error of the SLADV-Hop algorithm is lower than that of the other three algorithms.

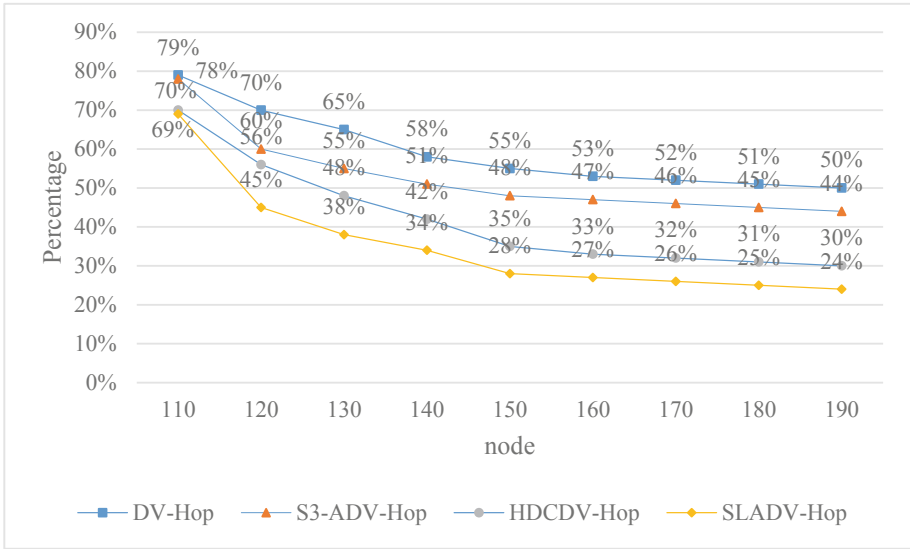


Fig. 2. The relationship between the total number of wireless sensor nodes and the positioning error

5 Conclusions

This article explores the application of UAVs in urban fire monitoring. After analyzing related theories, the UAV fire monitoring platform is designed. In order to improve the performance of the monitoring platform, the platform's positioning algorithm is optimized, and the optimized algorithm is tested, and the test results show that the positioning error of the optimized algorithm is lower than other traditional algorithms, which shows that the optimized algorithm is effective.

References

- Oliveira, M., Slezakova, K., Alves, M.J., et al.: Polycyclic aromatic hydrocarbons at fire stations: firefighters' exposure monitoring and biomonitoring, and assessment of the contribution to total internal dose. *J. Hazard. Mater.* **323**(pt.A), 184–194 (2016)
- Yu, Z., Zhang, Y., Jiang, B., Yu, X.: Fault-tolerant time-varying elliptical formation control of multiple fixed-wing UAVs for cooperative forest fire monitoring. *J. Intell. Rob. Syst.* **101**(3), 1–15 (2021)
- Parto, F., Saradjian, M., Homayouni, S.: MODIS brightness temperature change-based forest fire monitoring. *J. Indian Soc. Rem. Sens.* **48**(1), 163–169 (2019)
- Santos, J., Romeiro, J., Assis, J., et al.: Potentials and limitations of remote fire monitoring in protected areas. *Sci. Total Environ.* **616**, 1347–1355 (2018)
- Yin, H., Guli, J., Jiang, L., Yu, T., Umuhoza, J., Li, X.: Monitoring fire regimes and assessing their driving factors in Central Asia. *J. Arid. Land* **13**(5), 500–515 (2021)
- Sudhakar, S., Vijayakumar, V., Kumar, C.S., et al.: Unmanned aerial vehicle (UAV) based forest fire detection and monitoring for reducing false alarms in forest-fires. *Comput. Commun.* **149**, 1–16 (2020)

7. Kordos, J.: Tests of new method of monitoring endogenous fire hazard in hard coal mines. *J. Sustain. Min.* **18**(3), 134–141 (2019)
8. Yuan, L., Thomas, R.A., Zhou, L.: Characterization of a mine fire using atmospheric monitoring system sensor data. *Min. Eng.* **69**(6), 57–62 (2017)
9. Elaine, S.: Remote starting and monitoring for Davey firefighter pumps. *Power Equipm. Austral.* **37**(6), 12–12 (2017)
10. Rowland, I., Litton, C.D., Thomas, R.A.: Evaluation of detection and response times of fire sensors using an atmospheric monitoring system. *Transactions* **340**(1), 104–112 (2016)
11. Beata, P.A., Jeffers, A.E., Kamat, V.R.: Real-time fire monitoring and visualization for the post-ignition fire state in a building. *Fire Technol.* **54**(4), 995–1027 (2018)
12. Song, H.S., Lee, Y.Y.: Forest fire monitoring system using a panoramic image based on multiple cameras and measuring of distance. *J. Korean Inst. Illuminat. Electr. Install. Eng.* **32**(3), 44–52 (2018)



Auto Tailgate Automatic Control System Based on Single Chip Microcomputer

Ying Zhao^(✉)

School of Intelligent Manufacturing, Changchun Sci-Tech University, Changchun 130600, Jilin, China

Abstract. With the development and maturity of automobile electronic technology, the automobile is gradually developing towards the direction of comfort and intelligence. The electric tailgate technology is widely used and respected in the process of automobile product design, production and sales by each main engine factory. The quality of the vehicle directly affects the driving safety of the automobile. A car tailgate controller based on KEAZ128 microcontroller is designed in this paper, which has the functions of suction and unlock of tailgate door lock, electric lifting tailgate door, CAN bus and LIN bus communication. The suction function of the tail door lock reduces the impact of the pawl and ratchet in the lock, prolongs the service life of the tail door lock, reduces the noise of closing the door, and improves the safety of the car tail door. The electric lifting function makes the lifting of the tailgate more stable and the anti-clamping function improves the safety of the use of the tailgate. CAN bus and LIN bus body controller CAN make the controller and communication functions, realizes the tail door system and information interaction, body controller based on KEAZ128 SCM sport-utility vehicle tail gate system controller, including the overall design, circuit design and working principle of the controller and the software design, realized the tail gate of electric hoist, prevent clip; Automatic suction and electric unlock of tail door lock; CAN and LIN bus communication function of tailgate controller. The designed controller can effectively improve the running safety of the vehicle, improve the convenience and safety of the tailgate operation.

Keywords: Automobile tailgate · Controller design · Automatic control · Anti clip device

1 Introduction

With the rapid development of science and technology, people's needs are also growing, and intelligence has become people's tireless pursuit [1]. The advent of electronic computers has enriched and simplified the world. Of course, the micro-processing system also emerged quietly and was widely applied in practice. MCU control was used to handle some transactions, which reduced labor time, improved labor efficiency and improved accuracy [2]. The appearance of MCU has made great changes in life and production. People's life is more and more inseparable from network and intelligence. In order to improve the reliability and convenience of car door locks, the combination

of electric unlock structure and mechanical unlock structure has gradually replaced the traditional pure mechanical unlock mechanism [3]. The share of sport-utility vehicles in the Chinese car market has risen sharply in recent years. Consumers' utilization rate of tailgate system is far higher than that of other models, and they put forward higher requirements on the reliability, convenience and intelligence of tailgate use [4]. There are three main types of tail door electric mechanism used in automobiles, the first type is the electric pole type. This type of construction excludes the tailgate gas spring construction. The electric pole is directly positioned on the body of the car, which is compact in structure, beautiful in appearance and does not take up the side of space. It is mainly used for medium and large models [5]. The second type is swing arm connecting rod type, this kind of structure to keep the tail gas spring door, installed in body side surround electric connecting rod to drive the tail gate, swing arm with gas spring combination to realize the tail door opening and closing, the lateral structure occupies a certain confining space, beautiful degree is electric poles type difference, advantage is that can adapt to the tail door size bigger and bigger weight [6], so much quality larger MPV models applied in the tail. The third type is the electrically driven hinge type. The structure also retains the tailgate gas spring. The traditional car tailgate process requires manual dismantling, which is cumbersome and unsafe. With the rapid development of science and technology in our country, the development direction of automobile has become more and more comfortable and intelligent. The complicated operation of traditional automobile tailgate can no longer meet people's needs, so the intelligent electric tailgate, which can give consideration to operability and safety, emerges as the times require. It will be the car tailgate into the intelligent category, to achieve the goal of saving time and effort.

2 Proposed Method

2.1 Principle of Single Chip Microcomputer

The MCU is mainly composed of three parts: arithmetic unit, controller and register [7]. The arithmetic unit is composed of ALU, accumulator and register. First, the accumulator and the register input two 8-bit source data to the ALU, then the ALU completes the logical operation of the source data [8], and finally stores the operation results in the register. The controller is composed of program counter, instruction register, instruction decoder, timing generator and operation controller. It is the organization that issues the command, and the controller issues action commands to each part. The arithmetic unit receives the command and performs the corresponding operation, and stores the result of the operation in the corresponding register. MCU is now closely related to our daily life [9].

2.2 Automatic Control System of Automobile Tail Door Based on Single Chip Microcomputer KEAZ128 Single Chip Microcomputer

In the tailgate system, different from the traditional gas-spring tailgate design, this controller can control the tailgate strut motor to achieve the tailgate electric lift, with the function of full anti-clamping; It can control the double motors in the tail door lock to

realize the function of automatic suction and electric unlock of the tail door lock. CAN send instructions to MCZ33903 sub-chip in the controller to realize the communication function of CAN bus and LIN bus. The results are shown in Table 1.

Table 1. Functions of KEAZ128 MCU

Module function classification	
Module classification	Describe
ARMCortex-M0+Kernel	Based on ARM's Cortex-M 32-bit MCU kernel, 1.77CoreMark@/MHz single-cycle access memory, 48 MHzCPU frequency
System	System integration module; Power management control module; Miscellaneous control module; Bit operation engine; Peripherals bridge; Guard dog
Memory	Golden 128 KB flash; Golden 16 KB SRAM
The clock	External crystal oscillator or resonator, low range: 31.25–39.0625 kHz; High range: 4–24 MHz. External square wave input clock. Internal reference clock, 31.25–39.0625 kHz oscillator; 1 kHz low power oscillator LPO. Frequency locked ring FLL range 40–50 MHz

2.3 The Electric Tailgate System

(1) Structure and Composition of the Electric Tailgate System

Figure 1 is the structural framework of the electric stanchion electric tailgate system. The electric tailgate system is composed of ECU, pole module, snap-lock module, switch unit, buzzer, anti-clamping strip and kick foot sensor.

The controller module (ECU) is the core of the electric tailgate system. It takes high performance processing chip as the core and consists of driver module, signal acquisition module, A/D conversion module and communication module. It has various control functions such as signal transmission and analysis between vehicles, signal acquisition of electric tailplate system, signal acquisition of anti-pinch point, signal output of electric pillar drive, and self-priming driving signal output of tailplate lock.

(2) Anti-clamping Rubber Strip for Electric Tailgate System

The electric support module has the built-in hall sensor and the hall anti-clamp function. Generally, the backpressure of the lock body at the bottom of the tailgate is between 60N and 100N. If the back pinch is too small and the back pinch is calibrated too large, the back pinch frequency will increase, which will easily cause injuries to passengers. In addition, the rear windshield of the tailgate is located on both sides of the rear windshield closer to the hinge, and the backpressure arms are smaller. This position requires greater counter - extrusion pressure to trigger counter - extrusion, which is likely to cause injury to passengers. The controller module (ECU) receives the signal and controls the pole to open in reverse. Since

the signal feedback is through the conductor contact and compression deformation, the installation position requires a higher sealing gap, otherwise it is easy to cause incorrect clamping.

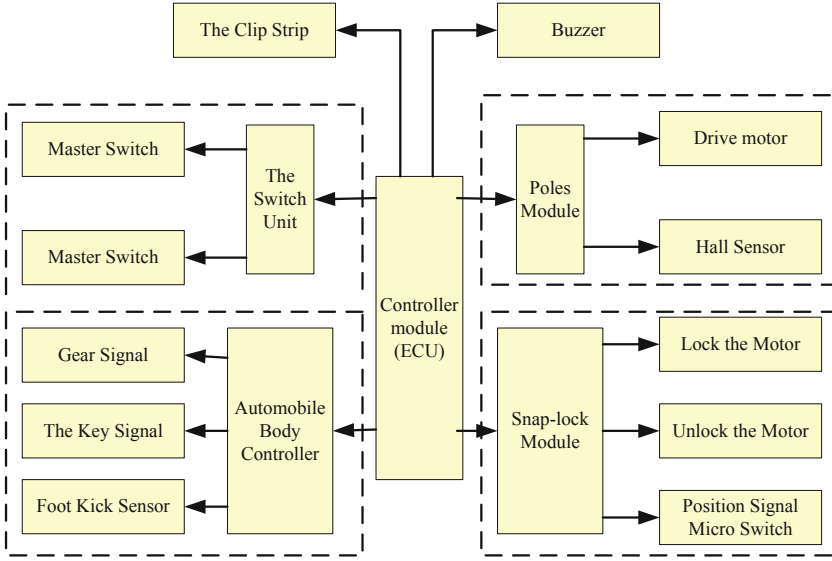


Fig. 1. Structural block diagram of the electric tailgate system

2.4 Signal Module of Auto Tailgate Automatic Control System Based on Single Chip Microcomputer

There are 4 signal lines in the tail lock, which are respectively used to judge whether the pawl position, the tail lock is in the semi-lock position, the unlocked position and the full lock position. When the signal is connected, it is equivalent to pressing a key, and the signal state is in a low level state. When the signal disconnects, it is equivalent to a key disconnecting, and the signal state is in a high level state.

According to the basic circuit design, the peripheral circuit of main chip and sub-chip is configured with 12 MHz crystal perk external crystal oscillator of the main chip, in which the main chip and sub-chip transmit instructions through SPI bus. 2.5 motor drive module motor drive module is an important part of the controller hardware circuit. In the design of the hardware circuit of motor drive module, two BTS7960 chips are used to form the h-bridge circuit to control the positive and negative rotation and speed regulation of the motor. The current diagnosis pin of BTS7960 is connected to the digital-to-analog conversion pin of the main chip, so that the controller can read the working current of the motor and judge the working state of the motor.

2.5 Closing/Opening Process of the Electric Tailgate System

The opening and closing of the tailgate is realized by controlling the positive and negative rotation of the strut motor through the motor drive circuit. In the process of closing the door, the controller module (ECU) first drives the pole to execute the closing action. During the execution, the hall sensor continuously collects the position signal and the anti-clamping signal. Stop the pole drive immediately when the reverse clamping signal occurs. When the pole returns to the open position or runs a certain distance in reverse, the system goes into standby again.

The sensor collects the position signal and the back clamping signal. When the reverse clamping signal appears, the pole immediately stops [10]. The system will go into standby again. When the action is in place, the system will open the door and enter the standby state. In the process of tailgate movement, when the motor speed or the mutation amount of motor current exceeds the set threshold, it can be determined that there is an obstacle blocking the tailgate movement, thus triggering the anti-clamping function, making the tailgate stop rotating or reverse rotating to achieve the anti-clamping function.

3 Experiments

3.1 Overall Design Scheme of Auto Tailgate Automatic Control System Based on Single Chip Microcomputer

The rear door system of SUV is mainly composed of the rear door strut, the rear door lock and the rear door, among which the movement of the rear door strut and the motor

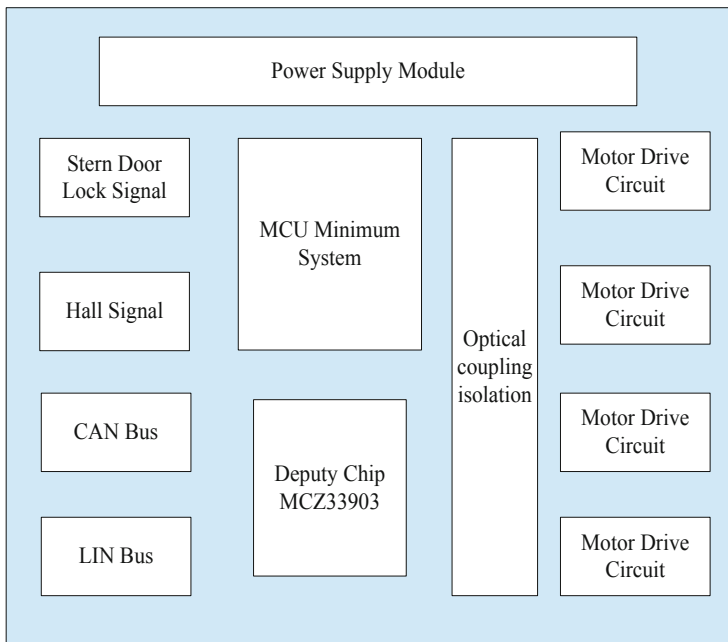


Fig. 2. Controller design scheme

state of the rear door lock are controlled by the controller. The overall design scheme of the controller is shown in Fig. 2.

3.2 Overall Software Design Scheme of the System

Through the analysis of the hardware circuit schematic diagram, it can be seen that the controller software design is mainly divided into three parts: suction and unlock of the tail door lock, opening and closing of the tail door, and CAN and LIN bus communication. The overall design flow chart is shown in Fig. 3.

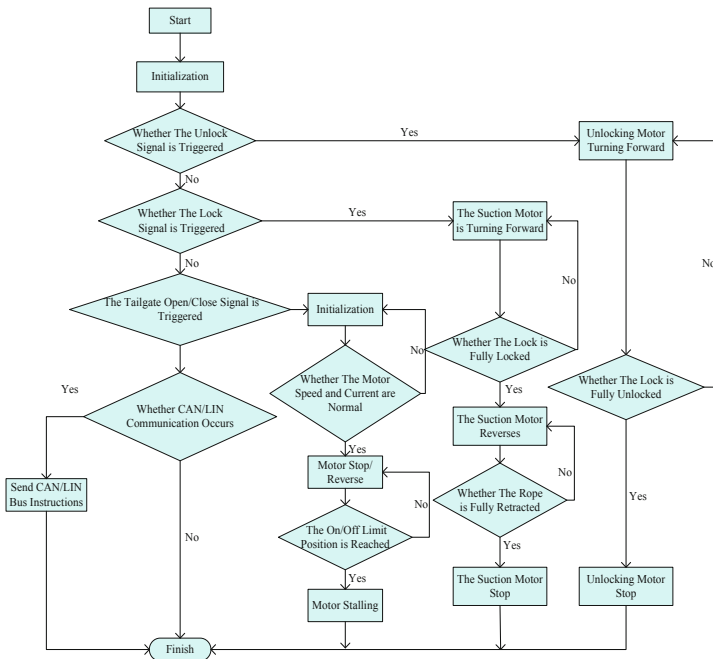


Fig. 3. Overall design flow chart

4 Discussion

4.1 Design, Calculation and Analysis of the Open and Close Operating Force of the Rear Vehicle in Manual Mode

(1) The Calculation and Analysis of Manual Starting Operation Force

During the manual opening of the tailgate, the manual operating force and the output force of the strut jointly overcome the gravity effect of the tailgate to realize the opening of the tailgate. The opening and closing of the tailgate are realized by controlling the positive and negative rotation of the strut motor through the motor

drive circuit. During the opening and closing of the tailgate, the load received by the strut motor will change. At this point, the resistance of the brace system f is opposite to the acting force of the assistant spring F_S , and the output force of the brace is $F_M = F_S - f$. According to the principle of torque balance, the operating force required for the manual opening of the tailgate at any opening degree F_{OPEN} can be obtained as follows:

$$F_{OPEN} = \frac{GL_G - 2X(F_S - f)L_S}{L_{OPEN}} \quad (1)$$

Where L_{OPEN} is the manual opening force arm with the hinge rotation axis as the pivot point. Through formula editing, the operating force curve of manual opening of the tailgate with the tailgate opening as a single variable at different slope angles can be output.

(2) Calculation and Analysis of Manual Closing Operation Force

The motor speed is obtained by using the hall signal. In the hall signal module, two hall signals need to work together. Hall signal 1 is used to count the number of pulses received within a certain time, so as to calculate the motor speed; When the hall signal 1 is in the rising or falling edge, the motor rotation direction can be determined by judging the high and low level of hall signal 2, so as to obtain the speed of the strut motor. In the process of tailgate movement, when the motor speed or the mutation amount of motor current exceeds the set threshold, it can be determined that there is an obstacle blocking the tailgate movement, thus triggering the anti-clamping function, making the tailgate stop rotating or reverse rotating to achieve the anti-clamping function. The manual operating force and the gravity of the tailgate jointly overcome the effect of the output force of the strut to achieve the tailgate closure. At this point, the resistance of the strut system f is in the same direction as the force of the assistant spring in F_S direction. The output force of the strut $F_M = F_S + f$ can be obtained according to the torque balance principle that the operating force required for the tailgate to close manually at any opening degree F_{CLOSE} is:

$$F_{CLOSE} = \frac{2X(F_S + f)L_S - GL_G}{L_{CLOSE}} \quad (2)$$

In the formula, F_{CLOSE} is the manual closing force arm with the hinge rotation axis as the pivot point. Through the formula editing function of excel, the manual closing operating force curve of the tailgate with the opening of the tailgate as a single variable is made.

4.2 Calculation and Analysis of Force Value of Electric Strut Mechanism

According to the high and low level states of the four door lock signals, the unlocking motor and the suction motor can perform three actions: forward rotation of the unlocking motor, forward rotation of the suction motor and reverse, so as to realize the electric unlocking and suction functions of the tail door lock. When the unlocking signal is sent, the unlocking motor will rotate until the tail door lock is in the unlocking state. When

the signal is sent, the unlocking motor will stop rotating. The unreasonable design of force value and elastic coefficient may lead to the failure of the opening interval of the tail door to maintain hover, and the problem of large operating force of manual opening and closing or failure of electric opening and closing will occur. When off signal, and the motor forward through the clues to pull the lock lever and lock movement, in the process of tail gate open and close, tail gate can stop at any position in the open and close angle and maintain, this state is called a hover, hover state support at the tail door by its own gravity and electric poles, the electric strut motor does not work, and the output force of the strut is the resultant force of the spring force of the boost spring and the resistance of the strut system.

5 Conclusions

Through theoretical analysis, the mechanical model of the opening and closing process of the electric support rod and the suction and locking process of the electric support rod are established, and the mechanical output curve of the function is edited by using the software formula. It provides a reference for the design of spring elasticity of electric prop, output force of driving motor and suction force of locking motor. In view of the current situation of designing tailgate system controller based on KEAZ128 automobile chip, a tailgate controller based on KEAZ128 microcontroller is designed in this paper, which has the functions of suction and unlock of tailgate door lock, electric lifting tailgate door, CAN bus and LIN bus communication. The suction function of the tail door lock reduces the impact of the pawl and ratchet in the lock, prolongs the service life of the tail door lock, reduces the noise of closing the door, and improves the safety of the car tail door. The electric lifting function makes the lifting of the tailgate more stable and the anti-clamping function improves the safety of the use of the tailgate. The functions of CAN bus and LIN bus enable the controller to communicate with the body controller and realize the information interaction between the tailgate system and the body controller. The sample of this scheme has been tested to meet the functional requirements.

References

1. Chunlei, J.: Design of deep-well thermometer and its compensation control system based on enhanced 51 single chip microcomputer. *J. Sens. Technol. Appl.* **8**(4), 123–128 (2020)
2. Zhang, Q., Zhao, J., Wen, H., et al.: Design of intelligent curtain control circuit based on single chip microcomputer. *Proc. Int. Conf. Artif. Life Robot.* **26**(3), 676–679 (2021)
3. Huirong, L.: Design of personnel-in-place system based on 51 single chip microcomputer and 4G module. *J. Sens. Technol. Appl.* **07**(4), 142–149 (2019)
4. Lopes, A., Guelton, K., Arcese, L., Leite, V., Bourahala, F.: Finsler-based sampled-data controller design for takagi-sugeno systems. *IFAC-PapersOnLine* **53**(2), 7965–7970 (2020)
5. Zhao, X., Zhang, X., Ye, X., Liu, Y.: Sliding mode controller design for supercavitating vehicles. *Ocean Eng.* **184**, 173–183 (2019)
6. Buljevic, A., Miletic, M., Mitrovic, A., Kapetina, M., Rapaic, M.: Examples of robust controller design. *Serbian J. Electr. Eng.* **17**(1), 65–82 (2020)
7. Ateunkeng, G.J., Alexandre, B.: Robust adaptive controller design for a robot manipulator. *Int. J. Contr. Autom.* **13**(4), 11–26 (2020)

8. Wang, H.W., Yu, X.C., Song, H.B., Lu, Z.H., Lloret, J., You, F.: A global optimal path planning and controller design algorithm for intelligent vehicles. *Mob. Netw. Appl.* **23**(5), 1165–1178 (2018)
9. Maslennikov, V.V., Meshcheryakov, V.V., Dovgopolaya, E.A.: Conditions for ensuring aperiodic transients in automatic control systems with a PID controller. *Open Autom. Contr. Syst. J.* **10**(1), 9–14 (2018)
10. Qiu, L., Wang, K., Wang, Z.: 69–78 GHz ESD-protected SiGe BICMOS PA with 30 dB automatic level control for mm-wave 5G applications. *Electron. Lett.* **54**(22), 1286–1288 (2018)



Corpus-Based Error Analysis of Students' Online Writings in Automated Essay Scoring System

Ronggen Zhang^(✉)

Department of Fundamental Teaching, Shanghai Publishing and Printing College, Shanghai, China

Abstract. Based on the corpora of 60 pieces of students' writings by the Contestant Candidates for National English Writing Contest for Technical and Vocational Students in recent years, the paper makes a comparative error analysis of expository writings by English Majors and Non-English Majors in a vocational college in Shanghai, utilizing tools such as PatCount, AntConc, and SPSS 19. Three main findings are arrived at as follows: first, there are significant differences between the expository writing by English major contestants and by their Non-English major counterparts, especially in the writing quality, where the former is obviously superior to the latter, despite that there were not apparent differences in their English proficiencies in the college entrance examination. Second, the non-English major contestants commit no errors in pronoun, which does not necessarily mean they are better at using pronoun than the English major contestants. Third, the most frequently committed errors by both the English majors and the non-English majors are those errors in sentence, wording, and word form. And the main causes for the occurrences of those errors are two factors, that is, the negative transfer of the mother tongue, and the carelessness of the learners. Finally, some pedagogical suggestions are made on how to improve the writing quality of the students.

Keywords: Corpus · Error analysis · Online learning · Automated essay scoring

1 Introduction

In 1960s scholars began to realize the importance of errors in English language learning. Since then error analysis has been applied as a useful method to analyze L2 learners' writing quality. According to Corder (1967), errors help teachers to know how well learners progress on the one hand; errors make learners realize that making mistakes is a means of their learning [1]. Ellis (1994) described five steps of error analysis suggested by Corder: collection of a sample, identification of errors, description of errors, explanation of errors, and evaluation of errors [2].

Sawalmeh found the Arabic speakers committed ten common errors, such as verb tense, word order, singular/plural form, subject-verb agreement, double negatives, spellings, capitalization, articles, sentence fragments and prepositions [3]. Kourosh and Mohsen analyzed writing errors in the genres of narration, description, and exposition,

and found the errors of noun form, verb tense, subject-verb agreement were the most frequent ones in expository writings [4]. Yang analyzed the causes of students' business English writing errors, and put forward the corresponding teaching strategies [5].

Sermsook noticed that English major Thai students commit most frequently committed errors such as punctuation, articles, subject-verb agreement, spelling, capitalization, and fragment [6]. Pescante-Malimas and Samson found the most common grammatical, syntactical, and mechanic errors included, grammatical errors like those in pronoun, tense, and subject-verb agreement, syntactic errors like fragments and run-ons, and mechanic errors like spelling [7]. Hussain found the common errors committed by Saudi female undergraduates were errors in grammar, lexis, semantics, and mechanics, and those in mechanics mainly included errors in punctuations, capitalization, and spelling [8]. Chaudhary and Zahrani concluded that the frequent errors occurred at the sentential level as well as the word level, where the spelling errors occurred most frequently while the demonstrative errors were the least [9].

The literature review from above reveals to us that researches on error analysis covers various subjects from college students to university undergraduates, either English majors or non-English majors. However, comparative study of errors analysis concerning both English majors and non-English majors is relatively rare to find. Hence, this paper attempts to fill in this gap by making the error analysis of expository writings by English majors and non-English majors in vocational college. The research process is done according to the five steps of error analysis suggested by Corder, described in Ellis (1994): collection of a sample, identification of errors, description of errors, explanation of errors, and evaluation of errors [2].

2 Methodology

1) Sampling

The corpora concerned are based on the 60 pieces of students' writings by the Contestant Candidates for National English Writing Contest for Technical and Vocational Students in recent 4 years and all the candidates are the English majors (of Business English) and non-English majors (of Liberal Arts, and of Technology) from Shanghai Publishing & Printing College.

2) All the writings were typed, tagged, and categorized into errors according to the error tagging system in CLEC. They are: word form (Fm), verb phrase (VP), noun phrase (NP), pronoun (Pr), adjective phrase (AP), adverb (Adv), preposition phrase (PP), conjunction (Cj), wording (Wd), collocation (CC), and sentence (Sn) [10]. All the writings were done online in the form of graph compositions.

3) Data processing: including scoring, concordancing, and editing, by using the software such as PatCount, AntConc, TreeTagger3, SPSS 20, etc. Each piece of writing is scored through the automated scoring system provided by <http://pigai.org/guest.php>, just for reference. For the sake of data processing through SPSS 20, the group of English majors is marked as 1 in value, while that of non-English majors is marked as 2 in value.

3 Data Analyses

From Table 1, several points can be seen as follows: first, there is quite a difference in writing proficiencies between the English majors and the non-English majors, where the former's mean score is 84 while the latter only 74 with a rather bigger standard deviation; second, both groups commit most frequently the errors of sentence, wording, and word forms; finally, on the whole English majors commit more errors than non-English majors in the corresponding writings, and especially non-English majors commit no errors in pronouns, which may partly be explained by the concordance hit result reveals that the English majors use pronouns more frequently than the non-English majors, with a frequency of 195 and 135 respectively. That is, the more pronouns the students use, more frequently they may commit errors in pronouns.

Table 1. Descriptive statistics

	English	Non-English	English	Non-English
	Mean	Mean	SD	SD
Score	83.03	74	4.87	7.09
Sn	2.53	2.87	2.3	3.08
Wd	1.87	1.53	1.66	1.74
Fm	1.8	1.33	3.07	2.43
PP	0.83	0.33	1.62	0.71
NP	0.73	0.23	1.14	0.5
Cj	0.6	0.13	0.72	0.35
Adv	0.43	0.13	1.5	0.43
Pr	0.4	0.1	0.81	0.31
AP	0.33	0.03	0.92	0.18
CC	0.33	0.03	1.06	0.18
VP	0.2	0	0.61	0

Besides, the T-test tells us that there is a highly significant difference between the English major group and the non-English major either in the writing quality or error frequency.

Table 2 indicates that the significant differences in error frequency between the two groups mostly occur in conjunction, noun phrase, and pronoun, which leads to their differences in writing quality.

Table 2. One-way ANOVA (between groups)

	Sum of squares	Mean square	F	Sig.
AP	0.82	0.82	1.73	0.19
Adv	2.40	2.40	2.10	0.15
CC	0.60	0.60	0.96	0.33
Cj	4.82	4.82	17.28	0.00
Fm	3.27	3.27	0.43	0.52
NP	3.75	3.75	4.81	0.03
PP	3.75	3.75	2.39	0.13
Pr	2.40	2.40	7.25	0.01
Sn	1.67	1.67	0.23	0.64
VP	0.07	0.07	0.24	0.63
Wd	1.67	1.67	0.58	0.45
Score	1224.02	1224.02	33.04	0.00

Table 3. Correlations between errors by English majors or by Non-English majors

Eng	VP\CC	.85**	Wd\PP	.59**	VP\PP	.45*	Pr\PP	.39*	NP\AP	.38*	Fm\scor	.46*
	Pr\Fm	.61**	CC\Sn	.55**	PP\Fm	.41*	VP\Sn	.39*	Adv\Wd	.37*		
Non-Eng	Sn\PP	.45*	Sn\Fm	.44*	PP\Cj	.44*						

**P < 0.01 *P < 0.05

Table 3 indicates the existence of the following correlations:

On the one hand, as for the English majors, first, the word form error is positively correlated with the score (.46*), which means the top contestants are liable to commit some word form errors; second, there are ten correlations between the errors committed by the English majors as follows: VP is positively with CC (.85**); Pr with Fm (.61**); Wd with PP (.59**); CC with Sn (.55**); VP with PP (.45*); PP with Fm (.41*); Pr with PP (.39*); VP with Sn (.39*); NP with AP with (.38*); and, Adv with Wd (.37*). The correlations indicate that those the English majors who frequently commit errors in verb phrase also make mistakes in collocations, preposition phrase, and sentence; those who often commit word form errors usually make mistakes in pronoun and preposition phrase; those who commit wording errors are likely to make mistakes in preposition phrase and adverb; and finally, they often commit errors in adjective phrase and in noun phrase at the same time.

On the other hand, as for the non-English majors in error frequency, here are only three correlations between the error frequency of the non-English majors: Sn is positively correlated with PP (.45*); Sn with Fm (.44*); and PP with Cj (.44*). The most obvious difference is that the score is almost not correlated with any type of errors by the non-English majors, and their pronoun error frequency is zero. As for the non-English majors, they frequently commit errors in sentence, word form, preposition phrase, and conjunction, probably at the same time.

4 Error Analysis and Its Pedagogic Implications

The data analysis from above reveals to us that both the English major contestants and the non-English major ones commit errors most frequently in sentence, wording, and word form.

4.1 Sentence Errors

Example 1: Mainly in the Central African Republic, food production has been drastically reduced due to natural disasters, which contribute <Sn> to a sharp rise of food price. (subject-verb disagreement, contribute → contributes)

Example 2: They hold their view for the reason that the condition of the original family can greatly influence the development of children. <Sn> Because a family with good conditions can give children <NP> more learning resources and growth conditions. (sentence fragment, because →, because)

Example 3: Countless people allover our country came <Sn> in this city day and night never stop, but population since 2012 to 2017 actually don't <Sn> have an <Wd> notable growth, aged elder are dying, even more young people are encouraged to live in shanghai <Fm >, (tense error, don't → didn't)

Example 4: In these years, mobile payment <Sn> gone up by about seven trillion. (sentence error, gone up → has gone up)

Example 5: At present, China's restaurants, supermarkets, individual merchants and even the food market <Sn> can use mobile payment, because of the popularity of credit cards. (sentence error, the food market <Sn> can use mobile payment → mobile payment can be used in the food market)

The above cases of sentence errors committed by the Chinese students mostly result from the negative transfer of their mother tongue Chinese, for in the Chinese sentence, the speaker is not required to pay attention to the agreements of subject-verb such as in tense, voice, number, and the like, which causes the corresponding sentence errors to happen in the first 4 examples. Also, active voice passive voice is more often used in Chinese, which may lead to the occurrence of sentence error in example 5.

4.2 Wording Errors

Example 6: From this we can boldly predict that the next few years <Sn> will also be in a balanced trend. However, those who accepted <Wd> treatment by the fifth year

were nearly 25% <Wd> more likely to survive than those who did not,... (accepted → received)

Example 7: If we can not take useful means <Wd >, we may not control this tend <Wd >,... (means → measures; tend → trend)

Example 8: What's more, most people tender <Wd> to consider that women do not need a job. They insist that the only thing women need to do is to take care of the whole family... (tender → tend)

Example 9: What's more, children can take control of their own destiny by making their life <Wd> choice in their life whether it is a big or a small one. (life → lifelong)

Example 14: For example, my younger female cousin, a <Wd> 8-year-old girl, always did <Sn> what she wants to do... (a → an)

Example 10: Therefore, countries in the development of economic and cultural <Wd> at the same time do not ignore the strengthening of public security capacity. (economic and cultural → economy and culture).

The causes of the concurrences of wording errors from examples 6 to example 10, may result from the negative transfer of the mother tongue Chinese of the students on the one hand, where they tend to translate literally the Chinese words like *jieshou*, *zuijin*, *cuosi*, and *yisheng* into the corresponding English words *accepted*, *last*, *means*, and *life*, in their writings. On the other hand, the occurrences of wording errors like *tend* (mistaking for *trend*), and *tender* (for *tend*) may be due to their carelessness. In example 9, the misuse of indefinite article *a* for *an* (*a 8-year-old girl*) may be both the result of the negative transfer of the mother tongue Chinese where are lack of such articles, and of the student carelessness. And in example 10, the misuse of *economic and cultural* for *economy and culture* reveals to us that the student has an inadequate knowledge of English parts of speech.

4.3 Word Form Errors

Example 11: In conclusion, men have more chance <Fm> in some key professions than women. (chance → chances)

Example 12: ...people can use the advanced equipment to solve these problem <Fm> and provide better growing conditions for crops. (problem → problems)

Example 13: The chart shows the growth of mobile payment scale between Chine <Fm> and the Unuted states <Fm> from 2012 to 2016. (Chine → China; Unuted → United)

Example 14: In addition, when using <Fm> mobile payment to pay, we will not have the problems such as change and counterfeit currency. (using → used)

Example 15: Conversely, the radio rate dropped continuous <Fm> <Adv> after reaching its summit 15% at about 1 p.m. (continuous → continuously)

The causes for the occurrences of above word form errors may be illustrated as follows: first, misuses of English plural noun forms lead to the corresponding errors from example 11 to example 12, and this may also be contributed to the negative transfer of the students' mother tongue Chinese, where there is lack of plural noun form use in modern Chinese. Second, those word form errors committed by the students from example 13 to example 14, may be purely due to their carelessness. Finally, the word form errors like

example 15 are mostly committed by the students owing to their inadequate grammatical knowledge, such as in the use of present participle, pre-modifier, adverb, and so forth.

5 Conclusion

From the above data analysis and error analysis, the following findings may be arrived at: first, there are significant differences between the expository writing by English major contestants and by their Non-English major counterparts, especially in the writing quality, where the former is obviously superior to the latter, despite that there were not apparent differences in their English proficiencies in the college entrance examination. The better writing performance of the English majors may result from their intense interest in English and also their hard work in the writing practice. Second, the non-English major contestants commit no errors in pronoun, which does not necessarily mean they are better at using pronoun than the English major contestants. In fact, the non-English majors tend to use fewer pronouns in their writings which may be due to the negative transfer of Chinese, whose native speakers prefer to use nouns repetitively rather than the pronouns. Third, the most frequently committed errors by both the English majors and the non-English majors are those errors in sentence, wording, and word form. And the main causes for the occurrences of those errors are two factors, that is, the negative transfer of the mother tongue, and the carelessness of the learners, despite there are also other reasons for them, such as the learners' inadequate knowledge of English grammar, and syntax, and also their less frequent writing practice.

Therefore, as for the instructors of those vocational college English, they need to instruct their students how to get rid of the negative transfer of their mother tongue, and how to improve their bad habit of carelessness in writing. More importantly, the instructors need to assign with the students more writing trainings, and urge them to read wider ranges of original English works.

References

1. Corder, S.P.: The significance of learners' errors. *Int. Rev. Appl. Linguist. Lang. Teach.* **5**, 161–170 (1967)
2. Ellis, R.: *The Study of Second Language Acquisition*. Oxford University Press, Oxford (1994)
3. Sawalmeh, M.H.M.: Error analysis of written English essays: the case of students of the preparatory year program in Saudi Arabia. *English Specif. Purpos. World* **14**(40), 3–17 (2013)
4. Kouroush, M., Mohsen, S.: The impact of text genre on Iranian intermediate EFL students' writing errors: an error analysis perspective. *Int. Educ. Stud.* **8**(3), 122–137 (2015)
5. Yang, Y.: Analysis of writing errors in business English writing for business-related majors: — taking Dalian Neusoft University of information as an example. In: *International Conference on Modern Education and Information Management (ICMEIM)*, pp. 120–123 (2020)
6. Sermsook, K., Liamnimitr, J., Pochakorn, R.: An analysis of errors in written English sentences: a case study of Thai EFL students. *Engl. Lang. Teach.* **10**(3), 101–110 (2017)
7. Pescante-Malimas, M.A., Samson, S.: Linguistic error analysis on students' thesis proposals. *IAFOR J. Lang. Learn.* **3**(2), 193–209 (2017)
8. Hussain, R.: An analysis of undergraduate Saudi EFL female students' errors in written English essays. *Arab World English J.* **1**, 241–258 (2019)

9. Chaudhary, A., Zahrani, S.A.: Error analysis in the written compositions of EFL students: a classroom study. *Int. J. English Linguist.* **10**(2), 357–366 (2020)
10. Yang, H.: Chinese learner English Corpus (CLEC). Shanghai Foreign Language Education Press, Shanghai (2002). (in Chinese)



Design and Development of Public Art Design System Based on VR Intelligence

Zhiyuan Niu and Shishi Xiang^(✉)

Nanjing University of Science and Technology, Nanjing, Jiangsu, China

Abstract. With the rapid development of modern society, VR intelligent technology has been widely used in various fields and has made great achievements. It can help designers and users solve various problems better and faster. At present, our country has not formed a systematic and standardized management mode for public art design. Therefore, it is necessary to transform the existing public space to meet the needs of people's lives. Therefore, it also requires designers to constantly update their concepts and innovative thinking, and at the same time, they must keep up with the times to accept new things and create new things that have their own characteristics and meet the needs of people. VR intelligent technology plays an important role in the future social development trend. It applies computer network technology to the real world and can better solve the problems of traditional art design. This article adopts experimental analysis and data analysis methods, and aims to take advantage of VR intelligence and apply it to the research and development process of the design system, so as to innovate the methods and expressions of artistic design work, and improve the quality and efficiency of work. According to the research results, the response time of the system is 36 ms, 41 ms, 59 ms and 78 ms respectively. The public art design system in this experiment has a shorter response time and higher accuracy rate, so it can meet the basic needs of design work.

Keywords: VR technology · Public art design · Visual image · System development

1 Introduction

With the advancement of science and technology and the improvement of people's living standards, the requirements for artistic design and creation are getting higher and higher. Traditional design methods can no longer meet the creative ability and aesthetic requirements of modern society for artists. At the same time, with the development of science and technology, VR intelligent technology has been widely used in various fields, and its application to public art design is conducive to providing more humane services to the public. Therefore, it is imperative to introduce VR smart technology in the field of public art works.

At present, many scholars have conducted research on VR intelligence and art design, and their research results are very rich. For example, Zhang Yu analyzed the overall

functions that a three-dimensional digital exhibition hall must have, and believes that the digital exhibition hall based on virtual reality technology is easy to use, highly interactive, and has practical value [1]. Tong Fang pointed out that with the industrialization of VR technology and the public's demand for informal education, museum institutions have adopted VR technology as a means of cultural dissemination [2]. Xu Zhili believes that the application of VR technology to art design can update the expression of art design, create a realistic virtual environment, and bring real experiences and feelings [3]. Therefore, this article is based on VR intelligence to conduct research on the design and development of public art design systems, which has important research significance and application value.

This article mainly discusses these aspects. First, a general introduction to VR technology is given. Then, it elaborates on public art design and related content. In addition, the application of three-dimensional Chinese characters in public art design is also introduced. Finally, an experimental study was carried out around the public art design system, and related experimental results and analysis conclusions were drawn.

2 VR Technology and Public Art Design

2.1 VR Technology

VR technology, also known as Virtual Reality. Studies have shown that the most important feature of VR technology is visual simulation, and later generations of visual simulation have also proposed various research methods. VR technology has the characteristics of immersion, interaction and design. It receives appropriate visual, auditory, tactile and even olfactory information through the device. When people's actions change, these feelings also change.

The ability of the human visual system to process 3D information can also be understood as the ability of humans to perceive 3D. The main principle of this ability is that the human's left and right eyes see different objects, and the brain calculates and processes the depth of the various scenes it sees to train humans' three-dimensional perception. This ability, also known as binocular parallax, provides a theoretical basis for people to experience a more realistic and immersive visual experience. The philosophy behind the virtual and reality is like a ruler, constantly measuring the minimum limit value of the virtual world and the real world for us. The first prototype of VR technology was developed from the exploration of human perception of the objective world. Although the theory at that time was far from perfect, the way of thinking they developed has sowed the hope that VR technology will formally create history [4, 5].

VR technology is based on computer networks and creates three-dimensional scenes through space simulation and virtualization. Allow users to more intuitively feel the sensory effects brought by artistic design in use. It loads images or video signals into the virtual environment through digital information transmission equipment (such as computers). Visually, you can see the shape changes of the surface of the object and the surrounding things and the scene; you can also see the connection between the inside and the outside of the object. You can simulate various objects, processes and results with different functions according to your needs, and realize people a variety of application modes such as the combination of computer interactive interface and multimedia system,

It mainly exhibits the three characteristics of immersion, interactivity and imagination. Immersion describes the fusion of the virtual world and reality under the multi-sensory sense of sight, hearing, touch and smell as soon as the user enters the virtual environment. Interactivity must first ensure the authenticity, efficiency and real-time nature of user operations, natural observation and operation of the virtual environment, and natural feedback. In the old technology center, most of the time, users ended input by clicking the mouse, typing on the keyboard, etc. Create a richer imagination in the virtual environment through the immersion and real-time interaction it creates. At the same time, the open association space of users expands the cognitive limit of human beings [6, 7]. The elements included in the work of the virtual reality system are shown in Fig. 1.

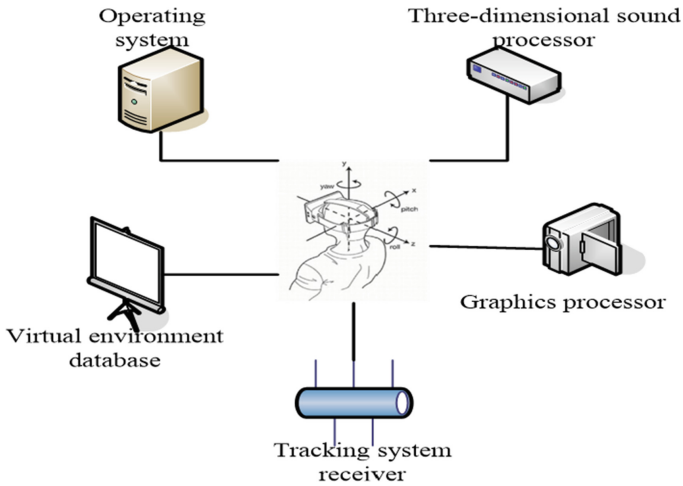


Fig. 1. Elements included in the work of a virtual reality system

2.2 Public Art Design

Whether in the West or in China, there is still no specific concept of public art. When historians describe the so-called public works of art in ancient history and culture, they mainly refer to certain forms of traditional art, but they are actually formed in public places. Various forms of art language are connected and separated from each other, forming the current public art.

The artistic design of public space is embodied in the promotion of space, the resonance and interaction of art, and its expressions are rich and diverse. First of all, the term “public art” belongs to the specific environment of “public space”. Whether it is urban outdoor spaces, landscape parks, commercial buildings, cultural alleys, urban pedestrian areas or indoor exhibition halls, light rail private rooms, museums, event centers, and large shopping malls are all public spaces. If a design or art form exists in a private space, it is not public art. However, under special circumstances, private spaces can be

transformed into public spaces, and private collections can be transformed into public art. For example, the palaces of ancient kings historically belonged to private rooms, but are now considered public tours. Secondly, public space art embodies the inheritance of people's material civilization, the promotion of traditional culture, is the embodiment of people's spiritual vision, and shows people's lifestyle and production methods. Therefore, under this kind of public art work that gives the soul of design, people will resonate and interact with it [8, 9].

The public sphere is a place of activity with people as the main body, and a social ideology with people as the main body. It is closely related to the ecological environment, natural environment, human race and social civilization. Through artistic design techniques, looking at the public environment from an artistic perspective is the best way to optimize the human living environment and its surrounding environment. Throughout history and social development, changes in the social environment are closely related to changes in the public space environment. At the same time, changes in the concept of a harmonious society will inevitably lead to changes in the public space environment. It has a certain degree of openness and at the same time it is open. Public space must be a space form that can be freely designed together and recognized by the public. Designing public art is a way to solve these problems while satisfying audience participation and appreciation. The influence of public space art on architectural design and decoration permeates many aspects such as art form, manufacturing process, expression theme, architectural exterior wall, space form and artistic technique, and the elements form a harmonious and beautiful whole. It uses artistic means to intuitively and comprehensively express the spatial form, color, texture, language and other art culture of the architectural theme. Today's architectural art has different characteristics in terms of colors, shapes and materials. It is not only satisfying functional needs, but also human aesthetic needs. It enriches the material world while also enriching people's spiritual world.

Public space can be subdivided into the following spatial levels: public space in the physical sense, public space in the social civilization level, and public space in the symbolic sense. Since the visual impact of art in public space is directly visible and difficult to change, advertisements for public art are not only reflected in the works placed in the public space, but also in the propaganda principles embodied in the works. According to Habermas's research, the concept of "public sphere" emerged with the social and historical development of western industrialized countries and the awakening of bourgeois consciousness. Public space is a communication space, everyone can freely enter and exit, and it is a place for the collection and communication of material, information and emotions [10, 11].

Therefore, these principles must be followed when performing artistic design in public spaces. The first is the principle of openness and fairness. The works of art in public space are initially accepted by the public, and it is not aimed at a specific group of people or free works of artists. Moreover, in many countries, public participation in public art is achieved through laws. The second is the people-oriented principle. The attributes of public art are different from pure art. Pure art emphasizes the expression of the artist's personal thoughts, the emotional expression and individualization of art, and the artist's personal aesthetic expression and emotional expression, while public art seeks

a foothold in order to preserve differences and integrate the public environment with art and the public One. The third is the principle of integrity. The relationship between public art and its environment is inseparable, and public art is not only the decoration of the environment. Public art works are no longer purely aesthetic needs, but must be realistic and coordinate with the overall needs of various facilities and spaces in real life.

2.3 Application of Three-Dimensional Chinese Characters in Public Art Design

In people's daily life, three-dimensional Chinese character design is very common, it can not only be used and appreciated as a two-dimensional graphic. And it can make people feel more intuitively that what the work wants to express is a visual form. Through this form, the public art design works can be made more complete, vividly showing the character modeling characteristics and environmental atmosphere, and can also make the whole work look more dynamic, giving people an immersive visual experience effect. In the case of two-dimensional Chinese characters, this is done when constructing the plane data of each stroke of the Chinese character. Whether it is lighting or texture mapping, the normal information from each vertex is used. The vertex normal information has been included in the surface structure of the three-dimensional feature. Therefore, the normal vector of each vertex of the three-dimensional Chinese character surface must be calculated and stored together with the spatial coordinate data of each vertex. The specific calculation method is shown in formulas (1) and (2).

$$M = E \times R \quad (1)$$

$$M_K = (M_1 + M_2 + M_3 + M_4) / 4 \quad (2)$$

Among them, E and R are the corresponding two vectors, M is the normal vector of the plane, K is the surface vertex, MK is the normal vector of the surface vertex, and M1, M2, M3, and M4 are the normal vectors of four adjacent planes.

3 Experiment and Research

3.1 Experimental Environment

In this experiment, the hardware configuration: CPU is Inter i5 3.1 GHz, memory is 16 GB, hard disk is 2 TB. Software configuration: operating system is Windows X, web server is Tomcat6.5, background database is MySQL5.8, development and debugging environment is Eclipse4.3. The network is configured with a bandwidth of wired 1000 Mbps.

3.2 Experimental Process

In this experiment, in order to better understand the function of the public art design experiment system, a total of 4 tests were carried out to test the system response time, CPU utilization, and Accuracy rate. Some test results are shown below.

4 Analysis and Discussion

In this experiment, a total of 4 tests were performed to test the system response time, CPU utilization, and Accuracy rate. The test results are shown in Table 1.

Table 1. Test result of system function

Testing frequency	Response time (ms)	CPU utilization (%)	Accuracy rate (%)
The first time	36	23	86
The second time	41	26	89
The third time	59	27	96
The fourth time	78	31	93

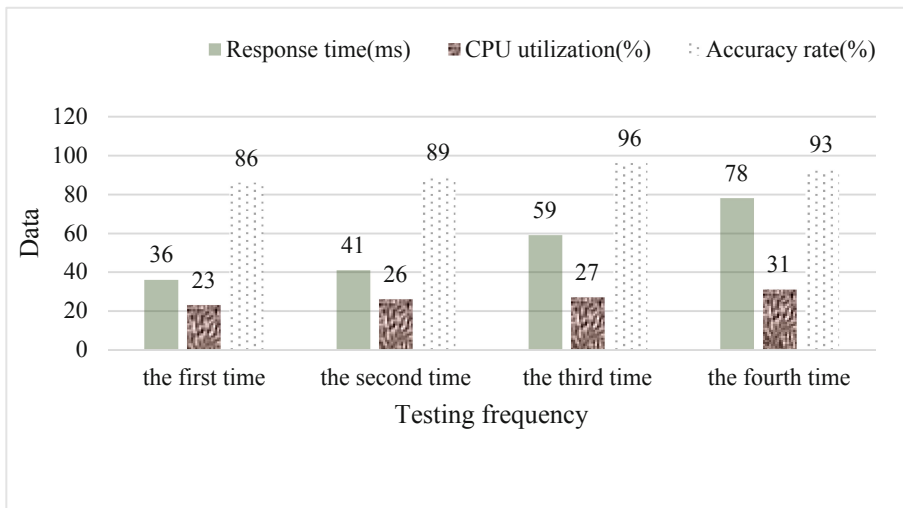


Fig. 2. Test result of system function

It can be seen from Fig. 2 that in this experiment, after the first, second, third, and fourth experiments, the response time of the system was 36 ms, 41 ms, 59 ms, and 78 ms, respectively, and the CPU utilization was 23%, 26%, 27% and 31%. It can be seen that the public art design system for this experiment has a shorter response time and a higher accuracy rate, so it can meet the basic needs of the design work.

5 Conclusion

With the continuous development of science and technology, people are paying more and more attention to concepts such as intelligence and personalization, which have also

had a certain degree of impact on traditional art forms. At the same time, with the rapid development and popularization of VR intelligent technology, it has been widely used in the field of modern design. VR smart technology is indispensable in the current social development and has potential and innovation. Applying it to modern design can improve work efficiency and reduce costs. Therefore, this article is based on VR intelligence and researches on the design and development of public art design system, which conforms to the trend and trend of the development of the times, and has important application value.

Acknowledgements. General Funding Project of China Postdoctoral Science Foundation (No.: 2020M670270).

References

1. Zhang, Y., Liu, Y.: Design and realization of 3D digital exhibition hall based on VR technology. (2016–3), 14–16 (2021)
2. Tong, F.: Coexistence of entertainment and education: cultural heritage VR design. (2017–6), 148–154 (2021)
3. Xu, Z.: The innovative performance of VR technology in art design. (2018–4), 60–61 (2021)
4. Lopez, J.P., Bosch-Baliarda, M., Martin, C.A., et al.: Design and development of sign language questionnaires based on video and web interfaces. *Artif. Intell. Eng. Des. Anal. Manuf.* **33**(4), 429–441 (2019)
5. Kara, E., Traquair, M., Simsek, M., et al.: Holistic design for deep learning-based discovery of tabular structures in datasheet images. *Eng. Appl. Artif. Intell.* **90**, 103551.1–103551.14 (2020)
6. Pal, S., Gupta, Y., Shukla, A., et al.: ActiveThief: model extraction using active learning and unannotated public data. *Proc. AAAI Conf. Artif. Intell.* **34**(1), 865–872 (2020)
7. Pan, Y., Meng, L., Li, Q., et al.: Research on interactive public art design methods in the context of information society. (2016–3), 34–37 (2021)
8. Fisichella, M.: Unified approach to retrospective event detection for event-based epidemic intelligence. *Int. J. Digit. Libr.* **22**(4), 339–364 (2021)
9. Ke, Z., Wen, Z., Xie, W., et al.: Group-wise dynamic dropout based on latent semantic variations. *Proc. AAAI Conf. Artif. Intell.* **34**(7), 11229–11236 (2020)
10. Murtaza, G., et al.: Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artif. Intell. Rev.* **53**(3), 1655–1720 (2019)
11. Kumar, K.: Design and development of travel and tourism recommendation system using web-scraped data positioned on artificial intelligence and machine learning. *Int. J. Adv. Trends Comput. Sci. Eng.* **9**(4), 5670–5679 (2020)



Research on Logistics Management Information System of Electronic Commerce Based on Computer Information Technology

Yang Ding¹(✉) and Neha Jain²

¹ Xi'an Peihua University, Xi'an 710125, Shaanxi, China

² Chitkara University, Punjab, India

Abstract. In order to do a good job in electric business logistics management in this article, taking the study of computer information technology, mainly analyzes the electricity business logistics management needs and current problems, after combining with the need to put forward the design scheme of logistics management information system, and introduces the application of system in logistics management process, application method, finally to verify this system, confirm the system is effective. Through this study, the logistics management information system can be constructed with the help of computer information technology. The system has been verified to be effective and can basically meet the general needs of e-commerce enterprises in logistics management.

Keywords: Computer information technology · E-commerce · Logistics management information system

1 Introduction

At present, China's rapid development momentum in the field of electricity, thus produced a large number of orders, logistics and logistics operations the internal situation has become more complex, the electric business enterprise to deal with all kinds of users demand, enterprise gradually discovered vulgar way of logistics management in the past no longer holds, not allows businesses to satisfy various customers' needs, Therefore, logistics management should be refined, personalized, self-service and other directions of reform. In this context, relevant fields believe that it is impossible to achieve the reform goal simply by relying on manual work or tradition, but computer information technology can break through the dilemma and achieve the reform goal. Therefore, it is necessary to carry out research around computer information technology, aiming to construct an effective e-commerce logistics management information system.

2 Needs and Current Problems of E-Commerce Logistics Management

2.1 Management Requirements

At present, the main management needs of e-commerce enterprises in logistics management are as follows: First, the whole process of supervision of logistics business

process, every link in the process should be supervised, so as to obtain the corresponding logistics business information, so as to understand the logistics situation according to the information, so as to make accurate management decisions; Second, it is necessary to improve the information exchange efficiency of logistics supervision, so as to avoid some of the more urgent problems can not be timely managed to cause more bad consequences; Third, it is necessary to improve the efficiency of logistics information and data processing, analysis and accuracy, namely by the electric with the current development situation of logistics management information, data level become very large, not only the type also more and more complex, lead to enterprise through artificial model to efficient processing of information and data, will lead to management decision irrelevant to the development of the realistic problems, It may also cause the asymmetry between management decisions and reality, so enterprises need to improve the efficiency and accuracy of logistics information and data processing and analysis [1–3].

2.2 Current Issues

Faced with various management needs, e-commerce enterprises have encountered some problems in logistics management, which originated from traditional logistics management methods and need to be paid attention to: First, in the logistics business process in the whole process supervision enterprises need to invest a lot of manpower, so you need to pay the corresponding labor costs, it's not all companies can undertake, at the same time, human will inevitably make mistakes in the work, can't cover, and the huge manpower injection, will bring a great burden on human resource management, thus trapped in a human model, Be caught in a dilemma; Second, in improving the efficiency of information interaction on the demand, the enterprise discovery relies on previous communications equipment to improve efficiency, such as the use of mobile phones can only achieve one-to-one communication, and may occur at the same time demand huge amounts of information interaction, mobile phones has limited the communication efficiency, which is enterprise facing a big problem; Thirdly, as mentioned in (1.1) above, the huge amount of information and data makes it impossible for human beings to process. Therefore, enterprises realize that the artificial mode needs to be reformed, but they have no good methods to achieve their goals, and it is difficult to get rid of the artificial mode, which is also a big problem faced by enterprises [4, 5].

3 E-Commerce Logistics Management Information System Design Scheme and Application Analysis

3.1 System Design Scheme

From the perspective of management needs, e-commerce enterprises have not found a good way to achieve various needs at present, which leads to related problems. In this case, computer information technology can provide enterprises with good help [6]. The design of e-commerce logistics management information system based on this technology can fully meet management needs and solve corresponding problems synchronously, so it is necessary to carry out system design. The design scheme of e-commerce logistics

management information system in this paper can be divided into four steps, as follows. Figure 1 shows the basic framework of e-commerce logistics management information system.

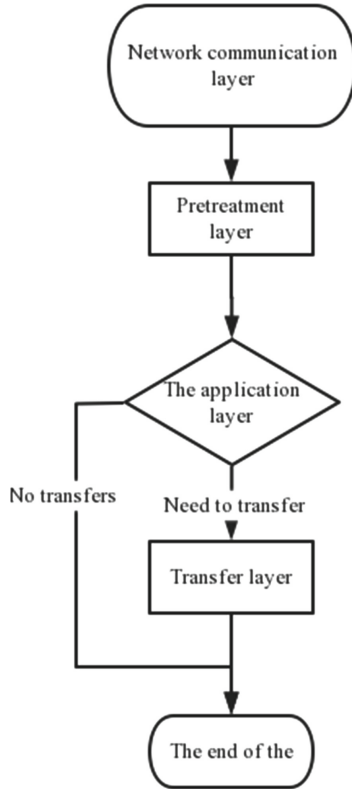


Fig. 1. Basic framework of e-commerce logistics management information system

3.1.1 Network Communication Layer Design

Information interaction is the necessary premise to realize electric business logistics management, the reason is that the logistics as a kind of practical activities, for the real physical time and space have certain dependence, but also restricted by physical time and space, electric business industry management departments can't be all the time in the field management, therefore, must establish the corresponding information communication channels, in order to ensure information can interact [7]. From this point of view, computer information technology can make the logistics management of e-commerce enterprises connected with the network, and the network is a good channel of information communication, and compared with the traditional way of information communication, the network has many advantages: First, the network can realize simultaneous communication online and offline message communication, on behalf of the interactive

information available at any time, without being limited by the physical time and space not only, also is not affected by the reality of human itself bring factors (online also can undertake information message each other even if they don't, know after such online information, to make management decisions); Second, China's modern network technology is relatively mature, which has greatly improved the communication efficiency. Therefore, using the network to communicate can effectively improve the efficiency of information and data interaction. On this basis, the design of the network communication layer only needs to configure computers and corresponding network hardware devices to connect to the network and realize network communication. However, it is necessary to pay attention to the configuration parameters of network equipment not to be too low to consider the needs of subsequent system operation. Table 1 is the basic configuration of the system in this paper.

Table 1. Parameters of basic configuration of the system in this paper

Indicators	Configuration parameters
The communication rate	54 Mbit/s
Data capacity	1 TB

In addition, because the network communication, information transmission on both ends for a network signals, so not only in the computer terminal to configure the network equipment, but also in the logistics business side configuration corresponding equipment, logistics business, for example, need to transport vehicles are installed on the sensor and transducer, it can use sensors to collect data and information, and the transmission of data information to the transducer, The transducer can convert the signal format data information of the sensor into digital format, and finally send it to the computer terminal through the Internet communication channel. On the contrary, the feedback information from the computer terminal will be received by the transducer, so that the field manual can consult, understand the management decision, and finally execute the relevant work.

3.1.2 Pretreatment Layer Design

Based on the network communication layer, the computer system may receive a large amount of information at the same time, but the initial state of these information is messy, and there may be useless, error and repeated information inside, so in order to facilitate the later reading, it needs to design a pretreatment layer. Rely mainly on big data pre-processing layer design and intelligent technology, namely the big data technology can provide powerful data support, and intelligence technology to rely on data support, help the computer system to identify the transmission of data, so that the computer system can classify information, and determine whether information error or is repeated problems, if there are similar problems, It will automatically delete. Through the pretreatment layer, the purity of information data can be improved and the magnitude can be reduced reasonably, which is conducive to the efficiency and accuracy of subsequent information application. In addition, the intelligent technology in the pretreatment layer mainly relies

on particle swarm optimization algorithm for data recognition. The algorithm is shown in Formula (1) and (2) [8–10]. Character pattern recognition can be developed through the two algorithms, and the recognition results can be compared with big data to know the information connotation, and then classify or perform other operations.

$$x_{ij} = x_{ij} + u_{ij} \tag{1}$$

$$u_{ij} = \omega u_{ij} + c_1 r_1 (p_{ij} - x_{ij}) + c_2 r_2 (p_{gj} - x_{ij}) \tag{2}$$

Where: $j = 1, 2, \dots, D$; C_1 and C_2 are learning factors. R_1 and r_2 are random numbers between $[0,1]$. P_{ij} is the individual optimal position of particle I , and P_{GJ} is the global optimal position.

3.1.3 Application Layer

This layer mainly integrates various logistics management functions for manual use. Therefore, this layer is directly related to whether e-commerce enterprises can comprehensively manage logistics information. The design task of the application layer is relatively simple, mainly for the design of application functions, among which the two most important functions are: First, data scheduling function, the function to determine whether a computer terminal can consult the logistics related business related information, and this information is the key to enterprises to make management decisions, such as through data scheduling functions of enterprises can understand the current logistics transportation vehicles, what kind of goods, every car transport the goods such as how to total value, It can also supervise the transportation of logistics vehicles with the help of sensors and other equipment. For example, in cold chain logistics, the temperature of vehicular ice storage can be monitored with the help of sensors. If the temperature is abnormal, it can remind the driver or the accompanying personnel, so as to play a management role. The key of data transfer function design lies in the database, that is, the database has the ability to store a large amount of data information, which is convenient for manual scheduling of relevant data. Secondly, some advanced databases have the function of automatic classification and storage, so as to avoid manual secondary sorting of data. In this paper, cloud database is selected. The database not only has large storage capacity, rich built-in function system, but also has data security. Second, data analysis function, which can analyze the transmitted data information, and the analysis will not only analyze the data information one by one, but also combine the data information (thanks to intelligent technology, not to mention here), so as to make more accurate management judgment according to the data information. For example, based on the vehicle-mounted ice storage temperature data and cargo refrigeration temperature standard data, we can know whether the current ice storage temperature meets the standard, whether the failure to meet the standard may lead to cargo decay, and how long the insurance can be, so as to make accurate management decisions. In addition, the application layer of the data analysis function also has a special transportation mechanism, namely if encountered in data analysis system to the events themselves have no legal power to make decisions, then it will be the input layer, data transmission and vice will put forward decision-making, on behalf of the management process, end will be the next cycle.

3.1.4 Transfer Layer

The main effect is to process the application layers can't make a management decision of data information and operation mode for the relevant data information, the application layer process of data analysis, analysis of the results, packaging, and then distributed to the human side, here, too, should with the aid of the network communication layer, such transfer layer can dock with the artificial mobile terminal, improve the convenience of the system. The traditional SQL database is selected in the design of the transit layer. The reason is that the transit layer only needs temporary data information storage, so there is no need to store large capacity database. The SELECTION of SQL database can reduce the cost of system construction. After that, the automatic distribution program is designed. After the distribution, the next cycle is carried out on behalf of the end of the management process. The program flow is shown in Fig. 2.

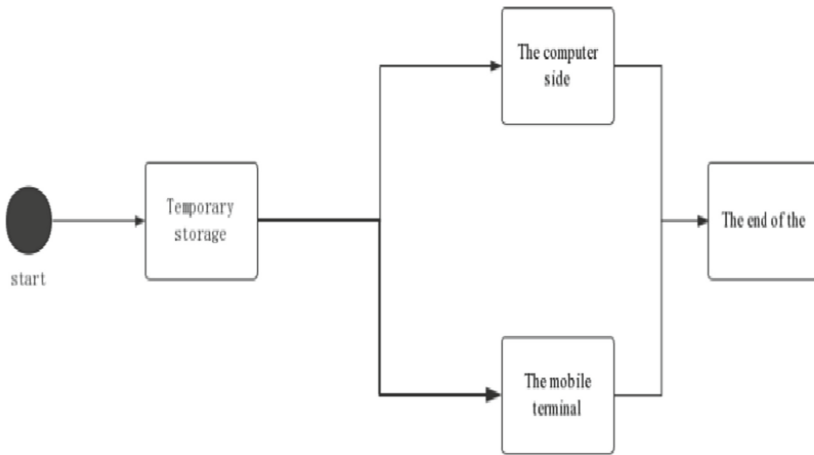


Fig. 2. Program flow

4 System Verification

4.1 Verification Scheme

The process of “goods out of warehouse → goods transportation → goods transportation completion” of cold chain logistics is the management goal, and the management work is carried out by relying on this system. During the process, the change of logistics status is simulated manually, such as changing the temperature of the vehicle ice storage. Through these methods, from the system response speed, system data information accuracy, system management decision correctness three indicators to judge whether the system is effective, whether it has practical value.

4.2 Verification Results

Table 2 shows the verification results of the system.

According to Table 2, the system response speed for the longest 3 s, so fast, that this system has strong real-time performance, and in the system data accuracy, information system management decision accuracy on the two indicators, the system performance is good, the same result showed that the system can provide accurate data information support, and make the right management decisions, Therefore, the system has applicability and can help e-commerce enterprises do a good job in logistics management.

Table 2. Verification results

Indicators	The results of
System response speed	1 s–3 s
Accuracy of system data information	100%
Correctness of system management decisions	100%

5 Conclusion

To sum up, with the development of e-commerce, logistics management is becoming more and more difficult, and labor has been overwhelmed, and traditional technical means are no longer applicable. From this point of view, computer information technology can help e-commerce enterprises break through difficulties and carry out logistics management work more efficiently, and the key lies in how to play the role of computer information technology, so enterprises should understand the nature of technology, carry out system design, with the help of the system can achieve goals.

References

1. Janssen, M.: The architecture and business value of a semi-cooperative, agent-based supply chain management system. *Electron. Commer. Res. Appl.* **4**(4), 315–328 (2006)
2. Rotchanakitumnuai, S., Speece, M.: Corporate customer perspectives on business value of Thai internet banking. *Soc. Ence Electr. Publish.* **5**, 270–286 (2004)
3. Chiara, V., Karen, V.: Managing risks in SMEs: a literature review and research agenda. *J. Technol. Manag. Innov.* **8**(3), 33–34 (2013)
4. Amar, L., Mu’Alem, A., Stößer, J.: The power of preemption in economic online markets. In: Altmann, J., Neumann, D., Fahringer, T. (eds.) *Grid Economics and Business Models*, pp. 41–57. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85485-2_4
5. Won, J., Ryoo, H., Kim, K.: The e-lollapalooza global workflow modeler: a registry-based workflow modeling methodology and system. In: *On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops; Lecture Notes in Computer Science*, vol. 3292. Collaboration Technology Research Lab. Department of Computer Science, Kyonggi University Suwonsi Kyonggido (2004)

6. P. Riquelme, I., Román, S.: Is the influence of privacy and security on online trust the same for all type of consumers? *Electron. Mark.* **24**(2), 135–149 (2014). <https://doi.org/10.1007/s12525-013-0145-3>
7. Ionescu, A.A.S., Serban, R.: Presentation of various types of electronic business available on the internet, advantages, disadvantages, key requirements and security, implementation model of an electronic business. *Int. J. Econ. Pract. Theor.* **2**(3), 179–196 (2012)
8. Hidayat, A.T., Dewantara, A.M.D., Saifullah, S.: The development of website on management information system for E-commerce and services. *J. Sisfok. (Sistem Informasi dan Komputer)* **9**(3), 380–386 (2020)
9. Halim, E., Karami, R.H.: Information systems, social media influencers and subjective norms impact to purchase intentions in E-commerce. In: 2020 International Conference on Information Management and Technology (ICIMTech) (2020)
10. Cahyo, M., Candiwan, C.: Analysis and design of sales information system on web-based E-commerce in yoga farm catfish breeding business using UML. *J. Media Inf. Budidarma* **4**(3), 683 (2020)



AHP-Based Small Cross-Border e-Commerce Export Trade Logistics Model Selection and Application Analysis

Jiang Yu(✉)

Beijing Institute of Technology, Zhuhai 519088, Guangdong, China

Abstract. With the advancement of economic integration and interconnection technology, the number of mobile Internet users has increased, driving the rapid growth of cross-border e-commerce (CBEC) in the era of economic information, making CBEC a new force in today's international export trade. This article analyzes the logistics status of a small CBEC company. The company mainly uses three logistics modes: dedicated line logistics, international express delivery, and fourth-party logistics, which account for 19%, 24%, and 28% respectively. Using the AHP method to analyze the most suitable solution for the export trade development needs of the company, it is concluded that the fourth-party logistics is the optimal solution. In the case that the fourth-party logistics cannot be delivered, the second-best solution is dedicated line logistics and finally international express.

Keywords: AHP analysis · Small cross-border e-Commerce · Export trade · Logistics model

1 Introduction

As e-commerce enters the international market, cross-border logistics models have also undergone various changes, and international export trade has increased significantly, which has greatly improved the development of cross-border logistics (CBL). Due to the highly complex logistics environment of CBL, high logistics costs and low efficiency, small CBEC companies need powerful technology to support logistics distribution. Choosing a suitable export trade logistics model to improve the level of CBL services is the research purpose of this article.

Many scholars at home and abroad have conducted research on the choice of AHP-based small CBEC export trade logistics mode, and have achieved certain results. For example, a scholar pointed out that there are various CBEC models, however, in view of the different development situations of CBEC companies, it is necessary to choose the logistics model with the least cost and the fastest efficiency in order to not only meet the logistics needs of customers, but also reduce the logistics costs of enterprises, and achieve a win-win goal [1]. A scholar believes that when small-value CBEC companies conduct international export trade, their logistics and transportation costs are high, and their

charging methods are complicated, which is not conducive to the e-commerce companies. Small-value CBL companies should meet their own development needs. Solve a series of problems in logistics transportation by establishing overseas warehouses [2]. Although the AHP-based small CBEC export trade logistics model selection research results are good, there are still many e-commerce companies that have not found a suitable logistics model. It is recommended that the AHP analysis method is combined with the actual situation of the company to choose the best one logistics model.

This article introduces several major logistics models for CBEC exports. Taking a small CBEC company as an example, an AHP model for logistics model selection is constructed based on the actual situation of the company, based on matrix diagram and AHP analysis. The availability judgment and consistency test of the logistics model indicators show that the best logistics model suitable for the development of the enterprise is the fourth-party logistics model.

2 Cross-Border e-Commerce Export Logistics Model and Index System

2.1 The Main Logistics Model of Cross-Border e-Commerce (B2C) Exports

(1) Postal parcel

As postal express accounts for a large proportion of my country's CBEC export trade business, international postal parcels have become the main body of trade export logistics. Suppliers in my country recognize that e-commerce companies have to control logistics costs even with meager profits, and can use postal parcels for commodities that are in urgent need of small items. In some remote areas and small countries, postal parcel delivery is the most economical. The postal parcel is suitable for small-scale and fragmented B2C retail e-commerce enterprises. The postal system provides a simple logistics service process [3]. Its advantages are simple charging methods, low transportation costs, and a wide distribution area, which can basically be delivered to corners of the world. The disadvantages are slow logistics, opaque logistics information, high package loss rate, and inability to transport large packages.

(2) International Express

International express delivery is a delivery method with high logistics operability in my country's cross-border B2C export trade. Generally, packages of objects with high commodity value and large volume and weight can be transported in this way. International express combines the advantages of wide delivery range, timeliness, low cost and traceability of logistics information. For example, using DHL's commercial courier service, it can be delivered to 220 countries and regions around the world, and it only takes 3–5 days for global delivery. The package logistics information can be checked at any time, and the package is not easy to lose. International express delivery is very suitable for customers who have higher requirements for delivery time, but most of the products exported by our country are in the middle and low-end, which is often difficult for CBEC to bear the high freight [4].

(3) Dedicated logistics

Dedicated line logistics mainly refers to the establishment of logistics warehouses in my country, and the special transportation mode of wholesale to international regions. The collection of parcels and goods is basically set up for regionally targeted special line distribution channels in most countries. In addition, this kind of logistics mode is simple to operate and requires no additional information when leaving the country, which is more convenient than the weight-based billing method [5, 6].

(4) Consolidated cargo flow

Collective cargo flow is a complete warehousing and distribution model developed based on dedicated line logistics. The cargo flow is consistent with the dedicated line logistics. The first step is to build a warehouse in China. The seller distributes the goods to the warehouse center built on the e-commerce platform. The buyer places an order on the e-commerce platform. The order will be shipped directly to the warehouse. Warehouse personnel pick up, pack, ship and other operations, and the goods are delivered to the destination country through a cooperative logistics provider, and finally to the consumer [7]. Consolidation logistics can use economies of scale to reduce transportation costs. Compared with postal parcels, this logistics model is safer, but it cannot be delivered to corner areas like postal parcels.

(5) The fourth-party logistics model

The fourth-party logistics model has been applied earlier in domestic e-commerce, and it has just emerged in the field of CBEC. The fourth-party logistics model of CBEC refers to that supply chain integrators integrate or reorganize resources from different international logistics companies, information technology, and match the logistics needs of buyers and sellers to reach logistics cooperation agreements, complete various transaction activities, and make progress. Export buyers and sellers, international logistics companies, provide one-stop complete and professional logistics consulting and services, that is, overall CBL service solutions. Fourth-party logistics companies often have abundant logistics service resources, strong supply chain integration capabilities, and great advantages in the quality and efficiency of CBEC logistics and distribution services. Due to the complex process of CBEC and its logistics activities, involving many buyers and sellers who belong to different borders, on the one hand, it provides a broad market for fourth-party supply chain providers, and on the other hand, it is engaged in CBEC of foreign trade companies provide more logistics options [8, 9].

(6) Cross-border logistics alliance model

With the advent of the "Internet +" logistics era, the logistics alliance cooperation model has also undergone new changes and developments. The most obvious representative form is the integration of traditional international commerce and CBEC models through CBL companies. Unify the whole, compared with the high investment brought by self-built overseas logistics distribution centers, how to cooperate with another CBL company or CBEC company will obtain a greater degree of capital savings and more convenient logistics service control. The current practice of CBL alliance is that several CBEC companies jointly invest in the construction and use of overseas distribution centers. CBL alliance members can transport their

goods to overseas warehouses before, and overseas consumers can purchase them. It is shipped and distributed directly from overseas warehouses, and then distributed to consumers by overseas logistics companies [10].

2.2 Evaluation and Analysis of Logistics Model Indicators

(1) Service availability

Service availability refers to the extent to which the logistics service of the logistics model meets the needs of customers, including the coverage of the logistics service, whether the product matches the logistics model provided, and so on. For CBEC operators, it is necessary to consider the matching of product weight and volume and other characteristics with the logistics model, and the scope of receipt and delivery of the logistics model. This dimension mainly considers product matching, weight restrictions and volume restrictions.

(2) Service reliability

Service reliability often has meanings such as security, assurance, and consistency. Product safety and on-time delivery are the basic requirements of logistics and distribution, as well as the requirements of meeting customers' consumer expectations and improving customer logistics experience. This dimension mainly considers safe transportation, safe distribution and accurate distribution [11].

(3) Service responsiveness

Service responsiveness refers to the degree of responsiveness of logistics services in the logistics model to customer needs. In CBEC transactions, most consumers hope to receive the purchased goods in advance, and merchants can solve logistics problems in a timely and efficient manner.

(4) Logistics cost

Logistics cost refers to the cost of logistics services provided by the logistics model. Considering the size of logistics costs is one of the important factors for enterprises to maximize their benefits. This dimension mainly considers transportation costs, distribution costs, storage costs, surcharges, and customs duties.

(5) Service flexibility

Service flexibility is the ability to describe the customer's ability of the logistics model to meet unpredictable logistics service needs. It mainly includes additional warehousing, transportation and distribution. This dimension mainly considers the ability to set up additional storage capacity, the ability to respond to special requirements, the ability to handle additional traffic and capacity, and the fulfillment of emergency or temporary orders [12].

2.3 AHP Analysis

Use YAAHP software to obtain the eigenvectors and eigenvalues of the judgment matrix, and standardize the eigenvectors, calculate the influence weight of factors of a certain level on the factors of the previous level, and then conduct consistency test. Calculate the maximum eigenvalue and one-time index of the judgment matrix:

$$\begin{cases} \zeta_{\max} = \sum_{i=1}^n \frac{W_i}{n \cdot W_i} \\ CI = \frac{\zeta_{\max} - n}{n - 1} \end{cases} \tag{1}$$

Among them, ζ_{\max} is the maximum eigenvalue of the judgment matrix, W is the eigenvector, and CI is a one-time index. Then calculate the consistency ratio:

$$CR = \frac{CI}{RI} \tag{2}$$

In the formula, CR is the consistency index, and RI is the average random consistency index. When $CR < 0.10$, the consistency of the judgment matrix is considered acceptable, otherwise the judgment matrix should be appropriately modified.

3 AHP-Based Small Cross-Border e-Commerce Export Trade Logistics Model Selection Research

3.1 Research Methods

(1) Matrix diagram method

On the premise of satisfying service availability, on the one hand, the company’s commodity customers have high requirements for service responsiveness, and on the other hand, the company pursues the purpose of maximizing profits. Therefore, it is necessary to make a trade-off between service responsiveness and logistics costs.

(2) AHP analysis method

International express delivery, dedicated line logistics and fourth-party logistics are relatively close in terms of service responsiveness and logistics costs, and it is difficult to qualitatively judge; thus, the five factors of availability, reliability, responsiveness, logistics cost and flexibility are integrated by using the analytic hierarchy process. Combine qualitative and quantitative selection.

3.2 Construction of AHP Model

- (1) Choose a plan from the three modes of fourth-party logistics, international express delivery and dedicated line logistics, so the target layer is marked as: T (optimal plan).
- (2) The factors that affect the choice of the optimal plan are availability, reliability, responsiveness, logistics cost and flexibility, so the criterion levels are marked as: A1 (availability), A2 (reliability), A3 (Responsiveness), A4 (logistics cost), A5 (flexibility).
- (3) The three alternatives are fourth-party logistics, international express and dedicated line logistics, and the scheme levels are marked as: F1 (dedicated line logistics), F2 (international express) and F3 (fourth party logistics). The AHP model selected by the company’s CBEC logistics model is shown in Fig. 1.

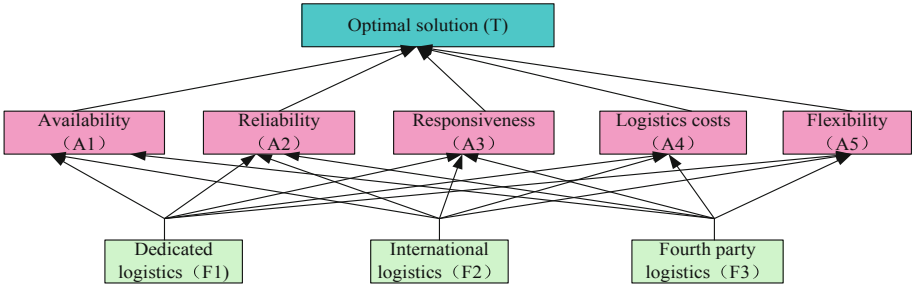


Fig. 1. AHP model

4 Analysis of Choice and Application of Small Cross-Border e-Commerce Export Trade Logistics Mode

4.1 A Company’s Cross-Border e-Commerce Logistics Model Sent to the International Market

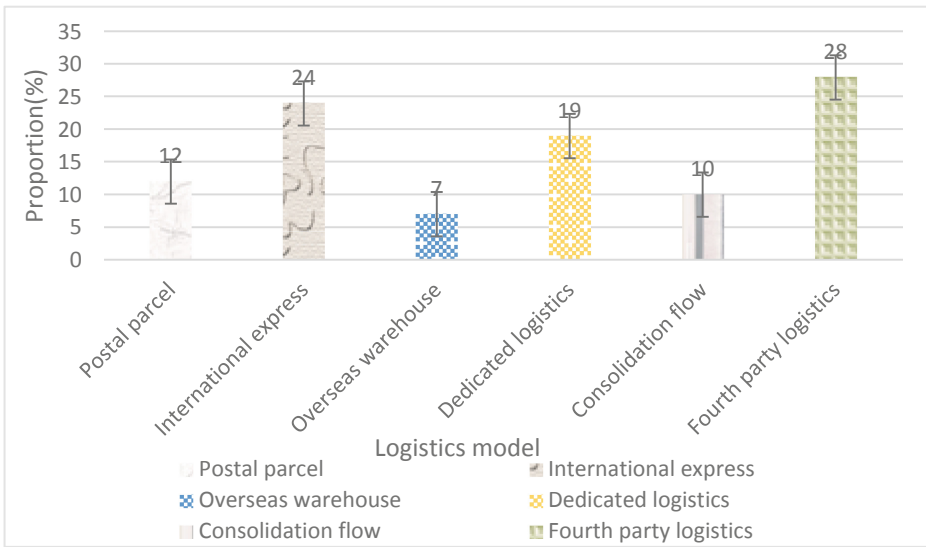


Fig. 2. A company’s CBEC logistics model sent to the international market

This article takes a company as an example to study the company’s CBEC logistics model to the international market. As shown in Fig. 2, the postal parcel logistics model accounted for 12%, the international express logistics model accounted for 24%, the overseas warehousing logistics model accounted for 7%, the special logistics model accounted for 19%, and the cargo logistics model accounted for 10%. The fourth-party logistics model accounts for 28%. It can be seen that the three logistics modes of international express delivery, special logistics and fourth-party logistics are the company’s preferred logistics methods.

4.2 Analysis of Logistics Mode Selection

Table 1. Consistency test of logistics model indicators

	F1	F2	F3
A1	0.045	0.02	0.071
A2	0.016	0.037	0.066
A3	0.033	0.028	0.052
A4	0.054	0.045	0.034
A5	0.069	0.036	0.068
Weights	0.217	0.166	0.291

As shown in Table 1, calculated according to YAAHP software: A1, A2, A3, A4, A5, the consistency index CR on F1, F2, F3 are all less than 0.1, that is, availability, reliability, The five logistics model indicators of responsiveness, logistics cost and flexibility have all passed the consistency test. The method of judging the optimal plan in the judgment matrix is to compare the weights. From the weight data in the table, we can see that F1 corresponds to a weight of 0.217, and F2 corresponds to the weight of is 0.166, the corresponding weight of F3 is 0.291, $F3 > F1 > F2$, therefore, F3 should be selected, that is, the company's logistics should choose the fourth-party logistics model.

5 Conclusion

This paper combines the actual logistics model of a small CBEC company, quantitatively analyzes the company's logistics model indicators, and calculates the weight of each indicator in different logistics modes through the analytic hierarchy process (AHP), and regards the largest weight as the enterprise logistics the best model, based on AHP analysis, finally concluded that the fourth-party logistics model is the best solution, and other logistics models can be used as alternative logistics models. It is hoped that the research in this article can provide reference opinions for CBEC companies at the same level when choosing a logistics model.

References

1. Koc, A., Turk, S., Şahin, G.: Multi-criteria of wind-solar site selection problem using a GIS-AHP-based approach with an application in Iğdir Province/Turkey. *Environ. Sci. Pollut. Res.* **26**(31), 32298–32310 (2019)
2. Meng, W.: Analysis of the competitiveness of cross-border e-commerce enterprises based on SPSS factor analysis. *Paper Asia* **34**(5), 44–48 (2018)
3. Zheng, J., Zhang, J.: Analysis on coordination mechanism of supply chain finance for B2C cross-border ecommerce. *Rev. Facult. Ingen.* **32**(14), 103–109 (2017)

4. Xiaojie, J.: Analysis on the pre-warehouse mode of fresh E-commerce market. *Progr. Soc. Sci.* **1**(2), 67–77 (2019)
5. Ye, Z., Huang, L.: Research on the influence of China's cross-border E-commerce tax reform policy on microeconomic subjects. *World Sci. Res. J.* **6**(3), 210–216 (2020)
6. Song, B., Yan, W., Zhang, T.: Cross-border e-commerce commodity risk assessment using text mining and fuzzy rule-based reasoning. *Adv. Eng. Inf.* **40**, 69–80 (2019)
7. Anson, J., Boffa, M., Helble, M.: Consumer arbitrage in cross-border e-commerce. *Rev. Int. Econ.* **27**(4), 1234–1251 (2019)
8. Shen, S.L., Lee, M.C., Lin, C.Y.: The influence of brand image and perceived value of cross-border e-commerce businesses on consumer purchase intention with Thailand's Baggo serving as the study object. *Int. J. Uncert. Innov. Res.* **2**(1), 57–71 (2020)
9. Gani, A.: The logistics performance effect in international trade. *Asian J. Ship. Logist.* **33**(4), 279–288 (2017)
10. Wei, H., Dong, M.: Import-export freight organization and optimization in the dry-port-based cross-border logistics network under the Belt and Road Initiative. *Comput. Indust. Eng.* **130**, 472–484 (2019)
11. Nita, H., Shah, X.: Three-layered integrated inventory model for deteriorating items with quadratic demand and two-level trade credit financing. *Int. J. Syst. Sci. Oper. Logist.* **4**(2), 85–91 (2017)
12. Regolo, J.: Manufacturing export diversification and regionalization of trade: which destinations for newly exported goods? *Int. Econ.* **151**, 26–47 (2017)



Design and Implementation of RFID-Based Logistics Clearance Information Platform

Qiong Yang^(✉)

Guangdong Mechanical and Electrical Polytechnic (GDMEP), Guangzhou
510400, Guangdong, China

Abstract. In recent years, with the rapid development of the Internet of Things and the deep integration of the Internet of Things and the logistics industry, logistics has gradually become a hot industry. The introduction of RFID technology into the logistics customs clearance process and the construction of an RFID-based logistics customs clearance information platform can simplify the work of administrative departments such as customs, taxation, inspection and quarantine, and improve work efficiency, thereby improving the efficiency of logistics customs clearance. Based on this, this article mainly studies the subject of “the design and realization of RFID-based logistics clearance information platform”. This paper designs the overall architecture of the logistics clearance information platform, which is divided into application layer, system interface, data platform layer and perception layer. This article combines RFID technology to design the core functions of the platform. In order to verify the effectiveness of the platform, it is found through testing that the average value of the platform’s page jump delay is about 1 s. This shows that the platform is stable, has good response performance and excellent parallel computing performance.

Keywords: Information platform · Logistics information system · RFID technology · Target recognition

1 Introduction

With the rapid development of the Internet of Things, the deep integration of the Internet of Things and the logistics industry [1–3]. In recent years, RFID has received widespread attention as a key technology for building “Internet of Things + Logistics” [3, 4]. The logistics customs clearance management introduces RFID technology, which can automatically and intelligently manage the data of each link of logistics customs clearance cargo information collection, management, container cargo tracking, and distribution [5, 6]. The application of RFID technology ensures the accuracy and effectiveness of data collection in all aspects of logistics customs clearance management, reduces the error rate and cost caused by traditional manual operations, and significantly improves the transparency and management efficiency of logistics customs clearance management. Therefore, about logistics customs clearance information Research on the platform has attracted a lot of attention [7, 8].

Regarding the research of logistics information, many scholars have conducted multi-angle discussions. For example, Zhou WH studied the anti-collision algorithm of RFID tags in logistics systems [9]; Kirch M's application of RFID technology in logistics and production mainly research on the application of intelligent logistics area [10]; Soong S takes medical logistics information as the research object, and studies the application of electronic paper and RFID/NFC in medical logistics information system [11]. It can be seen that this article takes the logistics customs clearance information platform as the research object, combined with RFID technology, can provide a certain reference for the development of logistics information.

This article first completes the overall architecture of the logistics customs clearance information platform, and divides it into application layer, system interface layer, data platform layer and perception layer. Then combined with RFID technology, a logistics customs clearance information platform for information exchange and sharing was designed. Finally, in order to verify the effectiveness of the platform, this paper conducted a test and concluded that the platform is stable, has good response performance and excellent parallel computing performance, and can be put into practical applications.

2 Design of RFID - Based Logistics Clearance Information Platform

2.1 Overall Platform Architecture

The entire logistics customs clearance information platform adopts a four-layer structure, including the common application layer, system interface layer, data platform layer and perception layer.

(1) Application layer

The application layer is responsible for all UI, including the functional modules of the platform, mainly including customs clearance data query, customs clearance services, port logistics management and other services. The application layer uses Internet of Things technology, big data technology, cloud processing and other technologies to realize the collection of logistics customs clearance data information. At the same time, the logistics customs clearance data will be retrieved from the data layer, and then the data will be summarized and sorted according to different modules, and finally the information Passed to the system interface layer.

(2) System interface layer

The logistics customs clearance information platform connects the logistics management systems of various ports, customs inspection, maritime affairs, and railway transportation systems through the system interface layer. The data interface layer is the bridge between the application layer and the data layer. The latter provides data support for the former, and the former provides functional demonstrations for the latter.

(3) Data platform layer

The Administration of Foreign Exchange, Entry-Exit Inspection and Quarantine Bureau, Customs, National Taxation Bureau and other institutions will also

upload data to the data layer through the open interface in the platform, and all data can be mined in the data layer. The data layer will make preliminary sorting and analysis of the information collected by the system interface layer, and store the preliminary sorted information in different small databases according to different types. Its core business runs through the use process of the entire logistics customs clearance information platform, from the shipper’s electronic cargo registration to customs clearance, to the dispatch and supervision of export cargo by the relevant transportation departments, all of which are inseparable from the data layer.

(4) Perception layer

The main function of the perception layer is to collect data for the information platform, relying on modern advanced information technology, using RFID, barcode, two-dimensional code, video, sensors, GIS/GPS, database, network communication technology and other types of information collection. The device collects data, completes the initial collection of data, and tracks the dynamics of customs clearance logistics in real time, and collects feedback data back to the platform in time.

In RFID technology, the reader adopts anti-collision technology to realize multi-tag identification of goods. Among them, this platform adopts the “Basic Unknown Tag Identification Protocol with Collision-Empty Slot Pairing with Collision-Empty Slot Pairing” (BUIP-CE).

The BUIP-CE algorithm consists of a known label inactivation phase and an unknown label collection phase. In each poll, the reader broadcasts two parameters, the frame length f and the random number r . After the tag is received, it uses the hash function to calculate the response time slot index through $H(ID||r) \bmod f$ [12].

The core idea of the BUIP-CE algorithm is to generate more readable time slots by adjusting the known tags from the expected collision time slot to the expected idle time slot, which improves the inactivation probability of known tags and speeds up the identification rate of unknown tags.

Assuming that there are n known tags in the RFID system, n_r represents the average number of inactivated tags in the original expected collision time slot and the new matching time slot, then the calculation method for the number of known tags inactivated in a time slot is as follows: (1) Shown:

$$E[d] = (n_i/f)e^{-n_i/f} \times (1 - 1/f)^{m_i} + [1 - e^{-n_i/f} - (n_i/f)e^{-n_i/f}] \times n_r \quad (1)$$

Finally, the probability of each known label being inactivated can be obtained by Formula (2).

$$P_d^i = (n_i \times f)/n_i \quad (2)$$

2.2 Platform Function Design

(1) Data query module

The main customers served by the platform include cargo owners, freight forwarders, supervision centers, customs departments, inspection and quarantine departments, railway departments, highway companies and financial companies. As

users of the system, users of these departments can obtain information on logistics clearance and transportation status in the system.

This functional module provides browsing and downloading services of policies and regulations, business declarations, applications, registrations, permits, filings and other materials in customs clearance business, and realizes centralized inspection of goods and regulatory information query services for platform users.

1) Centralized inspection sub-module

The centralized inspection sub-module mainly meets the requirements of “one customs and three inspections” for terminal container inspection, customs release and supervision. Its service objects are “one customs and three inspections”, shippers, freight forwarders, terminals, etc. The platform is divided into bayonet inspection, here is the comparison of the bill of lading data, if it is a one-time customs clearance, you can compare the customs declaration data, if it is a batch of customs clearance, for the barcoded documents, the bayonet inspection and release of the armed police can be used without input documents, but the electronic data of paper documents and delivery documents can be compared. The RFID tag information affixed to the goods is stored on the platform, which can realize automatic bayonet inspection and release. Only when the data does not match, will manual inspection and release be started.

2) Supervisory information query sub-module

The supervisory information query sub-module also provides container logistics statistics query functions, including cargo tracking, port aviation and service functions. The statistical query function of this module can be used by customs officers and enterprises. The difference is that enterprises can only query and count their own data. You can customize the fields and related constraints for querying statistics. Customs officers are divided into two levels. The general level can be used to view all the functional enterprises and the documents approved by the customs system. The advanced query can make statistics on the workload of the customs officers.

(2) Customs clearance service function module

The platform uses data exchange interfaces with customs, national inspections, maritime affairs and other departments to realize integrated functions such as port declaration, customs clearance filing declaration, and ship electronic joint inspection.

The platform provides self-service port declaration for the public and enterprises, and realizes the electronic declaration of entry and exit ships, the online declaration of import and export goods, and the remote declaration of passengers (crew) entry and exit. The various departments connected by the platform can share information, so these services only require the company to enter customs clearance data once, which helps to fully realize the one-stop customs declaration service and improve the efficiency of departmental collaboration.

3 RFID - Based Logistics Clearance Information Platform Test

3.1 Test Design

- (1) Customs clearance Matlab performs simulation simulation on the platform, and uses black box testing to find the following errors: incorrect or missing functions, interface errors, input and output errors, performance errors, database access errors, etc.
- (2) Record the test results after the system test: NG means that the error is large, NT means that the test cannot be run, OK means that the test is passed and run correctly, and POK means that most of the tests are correct.

3.2 Test Environment

The test environment of the logistics customs clearance information platform adopts the environment deployed in the enterprise intranet, and the hardware and software environment is shown in Table 1.

Table 1. Software and hardware test environment information table

Server nickname	Configuration information
Application server	CPU2 core 2.3 GHz; RAM 8 G; hard disk 300 G; system LinuxRHEL4.4.7-364 bit
Data server	PU2 core 2.3 GHz; RAM 8 G; hard disk 300 G; system LinuxRHEL4.4.7-364 bit
Client 1	CPU single core 2.3 GHz; RAM 8 G; hard disk 256 G; Windows 10 system; Firefox browser
Client 2	CPU single core 2.3 GHz; memory 4 G; hard disk 256 G; Windows 8 system; Google browser

3.3 Page Jump Delay Test

After the platform is online, user experience is very important. Therefore, it is very important that the platform can still run smoothly under different concurrent users. Based on this, a performance test was conducted. This test focused on the page jump delay time of the platform when the number of concurrent users was 50 to 150 respectively, and used the Jmeter test tool to analyze the performance of the server. During the test, different concurrent users were tested 8 times, and the minimum, maximum, and average page jump delay time were recorded.

4 Platform Test Results and Analysis

4.1 System Function Test Results

After testing, a small number of error codes and error data appeared on the logistics customs clearance information platform. The statistical results of the system test are shown in Table 2: The total test cases are 203 times, 89.66% of the test cases are correct, and 10.34% have correctable minor errors and display errors.

Table 2. System function test results

Total test case	Total number of tests	Percentage (%)
Actual test case	203	100
Total test case	203	100
NG	0	0
NT	0	0
POK	21	10.34
OK	182	89.66

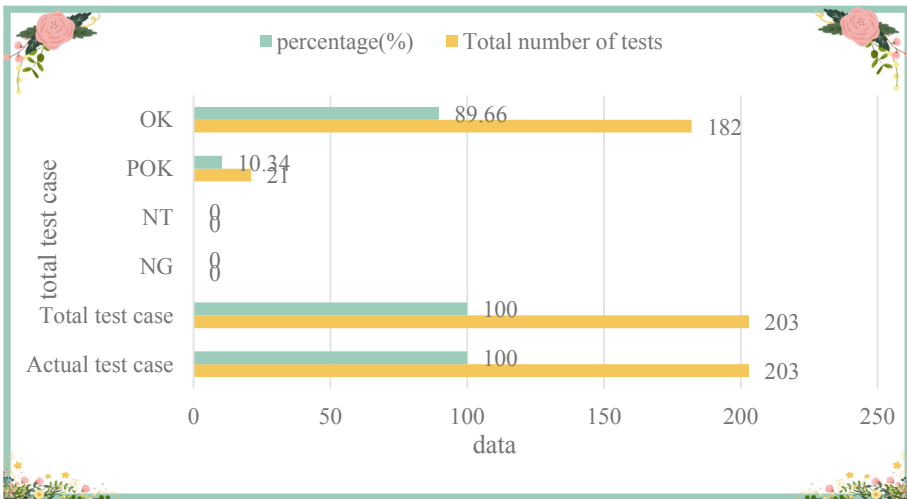


Fig. 1. System function test results

It can be seen from Fig. 1 that the construction of an RFID-based logistics clearance information platform basically meets the demand. At the same time, some incomplete test results appeared in the test, and the reasons were grammatical errors or small omissions. After modifying the program, the test results are all normal.

4.2 Page Jump Delay Test Results

Table 3. Page jump delay test results (s)

Number of concurrent users	Minimum	Average	Maximum
50	0.24	0.41	1.04
60	0.41	0.84	0.91
70	0.73	1.13	1.37
80	0.84	1.20	1.34
90	0.87	1.27	1.41
100	0.91	1.16	1.29
120	0.94	1.24	1.36
150	0.96	1.34	1.41
200	0.91	1.35	1.40
250	1.04	1.41	1.43

The page jump delay refers to the response time after the user clicks on the system page. This is also the main focus of system performance testing. The size of the system page jump delay largely determines the user's experience of the system. In the case of different numbers of concurrent users, the statistical results after the page jump delay test are shown in Table 3: When the number of concurrent users is 50, the average time of page jump delay is 0.41 s; when the number of concurrent users is 100, the page jumps The average transfer delay time is 1.16 s; when the number of concurrent users is 150, the average page jump delay is 1.34 s; when the number of concurrent users is 200, the average page jump delay is 1.35 s.

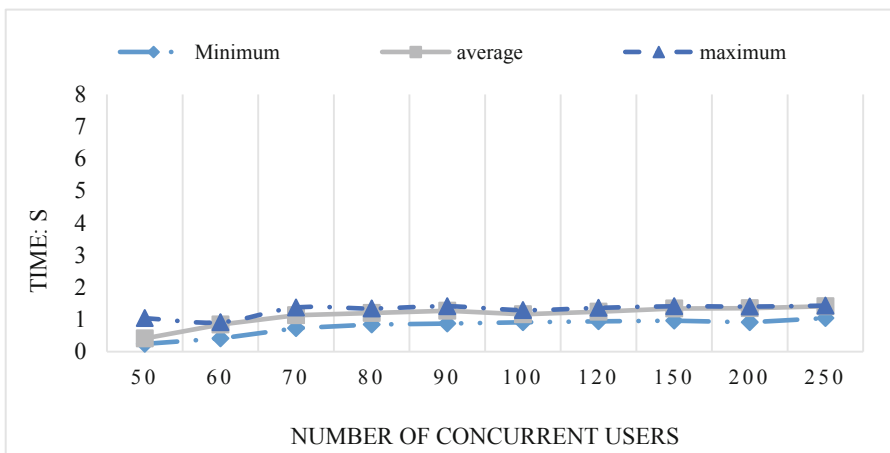


Fig. 2. Page jump delay test results

It can be seen from Fig. 2 that the average page jump delay of the platform is about 1 s, and the overall page jump delay of the platform is within the acceptable range of users, which meets actual application requirements.

5 Conclusions

Logistics customs clearance involves diverse and massive data information, including business data of enterprises, foreign exchange management, inspection and quarantine, ports, customs and other units and departments. Because these data have the characteristics of diversity, heterogeneity, open loop, standard inconsistency, etc., it has brought great inconvenience to the sharing of business information between various departments and enterprises. Therefore, the introduction of RFID technology to build a logistics customs clearance information platform for information exchange and resource sharing is conducive to improving the efficient operation of the logistics customs clearance business. After research, this paper designs a logistics information platform based on RFID technology, and the platform has a good operating effect, which can meet the requirements of practical applications.

References

1. Zhang, H., Jin, Y.: Credit system of smart logistics public information platform based on improved neural network. *Neural Comput. Appl.* **33**(9), 3987–4000 (2021)
2. Tu, M., Ming, K.L., Yang, M.F.: IoT-based production logistics and supply chain system – Part 2. *Ind. Manag. Data Syst.* **118**(1), 96–125 (2018)
3. Chen, C.-L., Deng, Y.-Y., Weng, W., Zhou, M., Sun, H.: A blockchain-based intelligent anti-switch package in tracing logistics system. *J. Supercomput.* **77**(7), 7791–7832 (2021)
4. Zheng, C., Zhao, X., Shen, J.: Research on location optimization of metro-based underground logistics system with Voronoi diagram. *IEEE Access* **8**, 34407–34417 (2020)
5. Li, J.: An effective system layout planning method for railway logistics centre in the background of big data. *Int. J. Reason. Based Intell. Syst.* **10**(1), 11–19 (2018)
6. Maldonado, S., Gonzalez-Ramirez, R.G., Quijada, F., et al.: Analytics meets port logistics: a decision support system for container stacking operations. *Decis. Supp. Syst.* **121**, 84–93 (2019)
7. Mardaneh, E., Loxton, R., Meka, S., et al.: A decision support system for grain harvesting, storage, and distribution logistics. *Knowl. Based Syst.* **223**(1), 107037 (2021)
8. Kong, X., Kang, K., Zhong, R.Y., et al.: Cyber physical system-enabled on-demand logistics trading. *Int. J. Prod. Econ.* **233**(7), 108005 (2020)
9. Zhou, W.H., Jiang, N.D., Yan, C.C.: Research on anti-collision algorithm of RFID tags in logistics system. *Procedia Comput. Sci.* **154**, 460–467 (2019)
10. Kirch, M., Poenicke, O., Richter, K.: RFID in Logistics and Production –Applications, Research and Visions for Smart Logistics Zones. *Procedia Eng.* **178**, 526–533 (2017)
11. Soong, S.: E-paper and RFID/NFC - a perfect partnership to revolutionise medical logistics. *Electr. World* **123**(1973), 22–23 (2017)
12. Mašek, J., Kolarovszki, P., Čamaj, J.: Application of RFID technology in railway transport services and logistics chains. *Procedia Eng.* **134**, 231–236 (2016)



Machine English Translation Scoring System Based on Deep Neural Network in the Internet of Things Environment

Longyuan Xiao^(✉)

Jiangxi College of Foreign Studies, Nanchang, China

Abstract. In machine translation, the evaluation method of the translation system is a very critical link, and it is also critical to the research of machine translation. The evaluation method helps to evaluate the characteristics of the translation system, and find the existing problems, so as to guide the development trend of the system. The correct evaluation method is the main cornerstone of the software system development process, and it is also the main force that promotes the development of machine translation systems. If there is no correct and effective evaluation method, it can be said that there will be no development of machine translation system. In this paper, we will study the deep neural network machine English translation scoring system under the Internet of Things environment, and understand the relevant theories of the machine translation scoring system based on the literature, and then design the machine English translation scoring system based on the deep neural network, and tested the designed system. The result of the test is that the score of the system in this paper is not related to the manual score, and the error range is within 10%.

Keywords: Deep learning · Neural network · Machine translation · English translation

1 Introduction

In the era of economic globalization, the links between countries in the world have become increasingly close, so of course the translation work between various languages is also very important [1, 2]. From the most basic travel abroad to the translation of international documents, to the cooperation of global companies such as global e-commerce, finance, and trade services, cross-language translation problems inevitably exist [3, 4]. However, how to efficiently, accurately and quickly translate text content into different languages through large-scale real-time changes to solve the urgent need for language translation is an urgent problem [5, 6]. Although manual translation technology can overcome this problem to a certain extent, the corresponding labor cost is also quite high, so the importance of machine translation is self-evident [7, 8].

Regarding the study of machine translation, some researchers pointed out that word-based automatic translation has limitations. When one word is translated into multiple

words, the word-by-word translation function does not work. Word translation is information that does not combine context, so it is difficult to capture general grammar. In addition, the word order adaptive model is relatively simple and cannot successfully capture complex word order adaptive information [9]. Some researchers have proposed a new model, a structure-based alignment model, which is a new idea to improve the alignment mode, which is mainly composed of two parts: coarse alignment and fine alignment [10]. Some researchers said that as deep learning has attracted increasing attention in the field of natural language processing, many researchers have begun to study it. However, unlike tasks such as image processing and language recognition, natural language processing often requires more complex model construction. For example, language modeling requires serialization, and parsing usually requires tree structure. Because natural language is an abstract entity of high-level knowledge created in the process of people's cognition, and sound and graphics are just primitive symbols for low-level entry, so although deep learning has made certain achievements in the field of natural language processing, the achievement are no high in graphics and sounds [11]. There are many research results on machine translation, but less research on machine translation quality scoring.

This article studies the machine English translation scoring system based on deep neural networks in the Internet of Things environment, and analyzes the limitations of machine English translation in the Internet of Things environment and the needs of the machine English translation scoring system on the basis of literature data, and then analyzes the machine English translation scoring system. The machine English translation scoring system of the deep neural network is designed, and then the designed system is tested, and the relevant conclusions are drawn through the test results.

2 Research on Machine English Translation

2.1 Limitations of Machine English Translation in the Internet of Things Environment

The translation system is mainly based on statistical methods to determine the meaning of a word. Because the semantics are flexible and changeable, it is usually determined by combining the grammar and semantics of the entire sentence, and it is easy to omit or make mistakes [12]. In addition, some semantics may be located at the beginning, middle, or end of the sentence. Even if the automatic translation system finds the correct meaning of the word, it is difficult to determine the position of the word in the translation. Moreover, online translation systems cannot learn independently. When encountering a word that does not exist in the main text, the system cannot request the network dictionary to get the correct translation, resulting in frequent translation shortages and incorrect translations. The automatic translation statistical system cannot perform common sense reasoning. Therefore, the translation system has defects in the subject, algorithm and translation model. In translation, prepositions and nouns, nouns (or other parts of speech, phrases or sentences equivalent to nouns) constitute preposition phrases. If the word collocation is different, the meaning of the preposition is also different. In most cases, when the translation system encounters a phrase with a more complex prosthesis or preposition, it cannot provide a correct translation.

2.2 Requirements of Machine English Translation Scoring System

- (1) Pursue accuracy. The basic goal of automatic evaluation is to be able to achieve evaluation capabilities similar to manual evaluation. Therefore, the main goal of automatic translation evaluation methods is always accuracy. In practice, the accuracy of automated valuation methods is generally reflected in the correlation between automated valuation results and manual valuation results. In recent years, the pursuit of accuracy and the simulation of manual evaluation by automatic evaluation methods are mainly reflected in the use of deeper language functions.
- (2) Pursue speed. Compared with the manual evaluation method, the most important aspect of the automatic evaluation method is its speed. As an important part of the system development cycle, the speed of the evaluation process directly affects the speed of system development. On the other hand, in today's most common statistical-based machine translation system, the automated evaluation method is not only a tool for the final evaluation of the system, but also a part of system debugging represented by minimizing error training. In this iterative training process the speed of self-evaluation is more important. However, the automated evaluation method based on deep language ability improves the accuracy of evaluation, but it often slows down the speed of evaluation. Therefore, in recent years, automatic evaluation methods based on string similarity have been widely used with obvious advantages such as fast utilization.
- (3) The pursuit of details. As the structure of machine translation systems becomes more and more complex, the limitations of traditional macro-evaluation-oriented automated evaluation methods have gradually emerged. The developers of the automatic translation system hope to get more information from the automatic evaluation process, that is, the evaluation information about the details of the automatic translation system.

2.3 Related Calculations

Replacing the original sentence with a phrase should affect the usability of the sentence. In order to ensure that the recently written sentences adapt to the grammatical habits of the language, this article introduces a language model with fluent characteristics. Simply put, a language model is a function that takes a sentence as input and returns the degree to which the sentence conforms to language habits. This feature allows the system not only to create words with the same semantics as the original sentence, but also to combine these words into a fluent English sentence. This model decodes the candidate translation set, that is, decodes the automatic translation result again, so this feature can re-evaluate the availability of the automatic translation result to a certain extent and select a better result. Based on the above characteristics, the type is defined as follows:

$$P(e) = \exp\left(\sum_{i=1}^4 \lambda_i p_i(e, f) + \lambda_{LM} LM(e)\right) \quad (1)$$

in,

$$P_1(e, f) = \varphi(f|e) \quad (2)$$

In the formula, h_i is each feature function, and λ_i is the weight corresponding to the feature.

3 Design of Machine English Translation Scoring System Based on Deep Neural Network

3.1 System Architecture

The system can evaluate the translation of texts provided by users and provide quantitative results of similarity calculations. Users can clearly and intuitively understand the translation quality based on the results provided by the system. According to the key characteristics of the system, the design system architecture model is shown in Fig. 1.

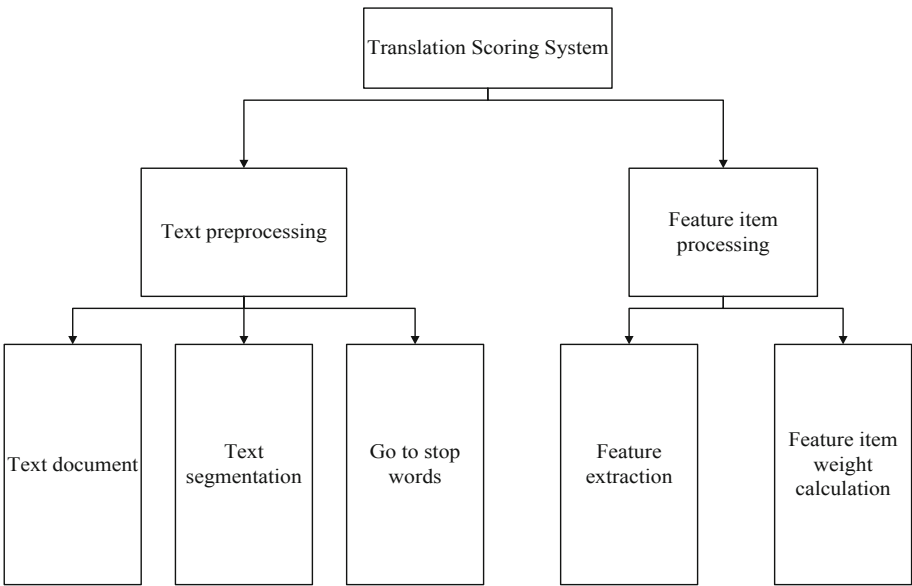


Fig. 1. System architecture model

3.2 Text Preprocessing

Before scoring the translation of the text, the text must first be preprocessed. Word segmentation is an important part of word processing. Because the accuracy of word segmentation directly affects the calculation of similarity. In this article, the NLPIR word segmentation system of the Chinese Academy of Sciences will be used to segment the text. The NLPIR word segmentation system is an open source word segmentation system developed by the Institute of Computer Science, Chinese Academy of Sciences. The system relies on hidden Markov multi-level model to achieve text segmentation. This

article uses this system to split the text into individual words and highlight the fragmented words as part of speech. Text preprocessing also requires that suspended words be deleted from the text after word segmentation. Some words that appear frequently in the text but have no meaning in the actual content and have little meaning for recognizing the content of the text are called stop words. For example, prepositions in text, these words often appear repeatedly in the text, but they have no practical meaning and cannot represent the theme of the text. The existence of stop words will cause large errors in the similarity calculation process, which can be considered as a kind of noise. Therefore, in order to improve the efficiency and accuracy of the calculation, the stop words in the text should be removed before the similarity calculation. The method of removing ending words in this article is to sort the ending words in the text according to the characteristics of ending words, create an ending word dictionary, and use the ending word dictionary to filter the text after word segmentation. If a word in the text appears in the terminal dictionary, it will be deleted until all ending words in the text are deleted.

3.3 Feature Extraction

The selection of attribute elements is a key component of the establishment of the text vector space model, and it directly affects the value of text similarity. For text with small content, although the text can be directly preprocessed, because the number of words in the text after word segmentation is small, the remaining various terms can be used as elements to estimate the similarity within the text. For large text, because the text contains many words. Regarding all preprocessed word elements in the text as text feature elements, it is difficult to determine the vector model of large text by calculating the weight value. This will greatly reduce the dimensionality calculation effect of vector modeling in text. Therefore, if you want to calculate the similarity rate of large text, you must perform more effective dimensional processing on vector text modeling, and remove less important or less resolution terms. If the word element appears more frequently in the text, its weight value will also be very high. Attribute elements often play a more critical role in the article, because they can be seen as similar to the content they represent in the article. Therefore, for larger content, the general method used in the article is to first count the weight value of each word element in the article, and then arrange the weight value of each word element from large to small. The higher the weight value of the vocabulary, when the article is used as an attribute element, the first 68% of the word elements in the weight are selected as the attribute element of the article, and this selected word element not only represents the majority of the article the content can also reduce the dimensionality of the vector model of the article.

3.4 Implementation of Deep Neural Network

Each subnet of the deep neural network receives a sentence from the sentence standby of the main body data as a representation of the fixed-length characteristics of the input. The two subnets share the same parameters. The model uses text pairs of the form (X_1, X_2) as input to the model. Among them (X_1, X_2) represents the text pair of the text data sample. X_1 is the word sequence of the first text, and T_1 is the number of words in the first text. X_2 is the word sequence of the second text, and T_2 is the number of

words in the second text. In order to match the sentence pairs of data samples in the same vector space, the deep neural network model mainly includes a word integration layer, a BiLSTM layer, an attention layer, a fully connected layer, and an output layer.

The model training process is the process of continuously updating all the parameters of the model until the model converges. If the initialization parameters are not configured correctly, the model will look inappropriate and overcorrect. The training process of the deep neural network model mainly includes the following aspects:

- (1) The model mainly uses the abandonment principle and the strategy of early termination to realize the solution, avoid the phenomenon of over-adjustment of the model and the disappearance of the slope, and accelerate the training speed of the model.
- (2) Use lateral entropy as a function of model loss and back propagation of stochastic gradient descent to train model parameters, and use Adam Optimizer to optimize the parameters of the entire model.
- (3) Use grid search strategy to find the best parameters for the model.
- (4) Use L2 fitting to optimize the loss function, and punish too much parameter weight during model training.

4 System Test

Table 1 shows the Pearson correlation coefficient between the experimental results of the Chinese-English data set and the translation quality value predicted by the system and the actual value. Among them, the Dev and Test columns represent the development set and test set respectively. Lines 2 to 4 correspond to the translated translation.

Table 1. Pearson correlation coefficient between the translation quality value predicted by the system and the actual value

Constraint	Predictive value	Test
1	0.557	0.526
2	0.521	0.509
3	0.533	0.517

It can be seen from Fig. 2 that the predicted value of translation quality is not much different from the true value of the system score, and the error is within 10%, indicating that the translation scoring system designed in this article is effective.

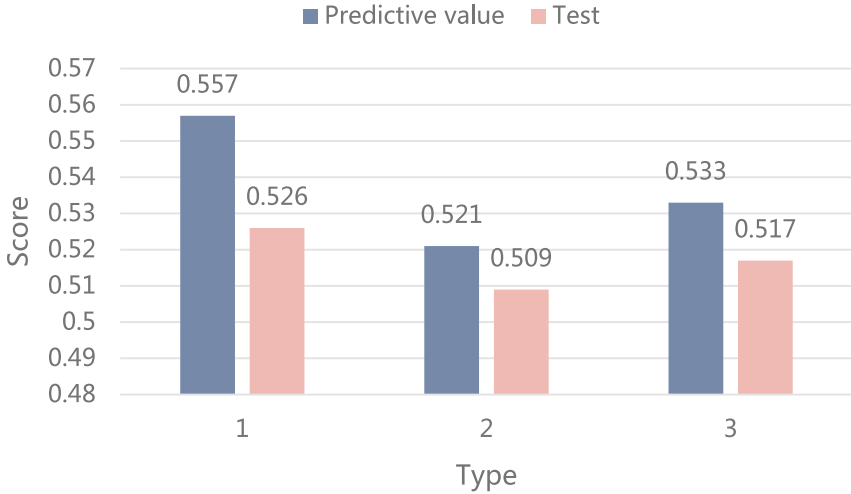


Fig. 2. Pearson correlation coefficient between the translation quality value predicted by the system and the actual value

5 Conclusions

This article focuses on the deep neural network machine English translation scoring system under the Internet of Things environment. After understanding the relevant theories, the machine English translation scoring system based on the deep neural network is designed, and the designed system is tested and tested. The results show that the predicted value of translation quality is not much different from the true value of the system score, and the error is within 10%, indicating that the translation scoring system designed in this paper is effective.

References

1. Rajeh, H., Li, Z., Ayedh, A.: A novel approach by injecting CCG Supertags into an Arabic–English factored translation machine. *Arab. J. Sci. Eng.* **41**(8), 3071–3080 (2016). <https://doi.org/10.1007/s13369-016-2075-9>
2. Chatterjee, N., Gupta, S.: Efficient phrase table pruning for Hindi to English machine translation through syntactic and marker-based filtering and hybrid similarity measurement. *Nat. Lang. Eng.* **25**(PT.1), 171–210 (2019)
3. Chakrawarti, R.K., Mishra, H., Bansal, P.: Review of machine translation techniques for idea of Hindi to English idiom translation. *Int. J. Comput. Intell. Res.* **13**(5(3)), 1059–1071 (2017)
4. Alsohybe, N.T., Dahan, N.A., Ba-Alwi, F.M.: Machine-translation history and evolution: survey for Arabic-English translations. *Br. J. Appl. Sci. Technol.* **23**(4), 1–1936124 (2017)
5. Koul, N., Manvi, S.S.: A proposed model for neural machine translation of Sanskrit into English. *Int. J. Inf. Technol.* **13**(1), 375–381 (2019). <https://doi.org/10.1007/s41870-019-00340-8>
6. Dahan, N.A., Ba-Alwi, F.M.: Extending a model for ontology-based Arabic-English machine translation. *Int. J. Artif. Intell. Appl.* **10**(01), 55–67 (2019)

7. Pathak, A., Pakray, P.: English-Mizo machine translation using neural and statistical approaches. *Neural Comput. Appl.* **31**(2), 1–17 (2019)
8. Lee, S.M., Briggs, N.: Effects of using machine translation to mediate the revision process of Korean university students' academic writing. *ReCALL* **33**(1), 1–16 (2020)
9. Novikova, A.V., Mylnikov, L.A.: Problems of machine translation of business texts from Russian into English. *Autom. Documentation Math. Linguist.* **51**(3), 159–169 (2017). <https://doi.org/10.3103/S0005105517030104>
10. Farzi, S., Faili, H.: Improving statistical machine translation using syntax-based learning-to-rank system. *Literary Linguist. Comput.* **32**(1), 80–100 (2017)
11. Haque, R., Penkale, S., Way, A.: TermFinder: log-likelihood comparison and phrase-based statistical machine translation models for bilingual terminology extraction. *Lang. Resour. Eval.* **52**(2), 365–400 (2018)
12. Chen, X., Gelernter, J., Zhang, H., et al.: Multi-lingual Geoparsing based on machine translation. *Future Gener. Comput. Syst.* **96**, 667–677 (2017)



System Design of Single On-line Uninterruptible Power Supply

Xifeng Shen¹, Ming Liu¹, Hua Liu^{1,3(✉)}, Jinfeng Dong^{1,3}, Guoqing Jiang¹,
and Xin Zhao²

¹ School of Engineering and Technology, Tianjin Agricultural University, Tianjin, China
tjwqns@tjau.edu.cn

² A&S Industry Technology (Tianjin) Co., Ltd., Tianjin, China

³ Hengshui Ruisida Machinery Factory, Hengshui, Hebei, China

Abstract. A high precision and output stability of the AC sine wave online uninterruptible power supply (UPS), the design is intended to demonstrate the advantages and disadvantages of the circuit design, get relevant data, so the output of 24 V switching power supply instead of the battery. In the selection of power supply circuit, the characteristics of one-way conduction of the transistor are used. When the 36 V power supply is supplied, the standby power side voltage is only 24 V, and the subsequent electricity is supplied by the power side. When the power side is cut off, the 24 V standby power supply (i.e. energy storage device) side becomes the high voltage side, and the transistor turns on to achieve the purpose of instantaneous switching. The inverter circuit based on the EGS002 inverter driver board and the power electronics converter technology, has the functions of stable output, low distortion rate, visual output data, etc., so that the uninterruptible power supply can output the single-phase sinusoidal alternating current with the frequency of 50 ± 0.2 Hz, $30 \text{ V} \pm 0.2$ V.

Keywords: UPS · Rectifier · Inverter · EGS002

1 Introduction

With the advancement of China's energy plan, the stability of the power grid has indeed been greatly improved, and power cuts have become extremely rare, but the user the power quality is still will be affected by some factors of force majeure or the influence of power use non-standard, causes there was a power equipment utilization or harmonic distortion, etc., seriously affect the user experience.

In addition, in 2013, the Eleventh Five-Year Plan for key projects, key projects for the development of the western region and the revitalization of the old industrial base in the Northeast, have expanded the market for UPS. In the UPS application, many STM32-based AC Sine wave online uninterruptible power supplies output corresponding DC power through isolation transformer, autotransformer and rectifier module, and access single-phase full-bridge inverter circuit through boost module to provide stable AC

voltage for the load [1]. Some instruments equipped with fault indication systems are equipped with UPS power modules. Compared with traditional fault indication system instruments, equipped with UPS power module fault indicator detection instruments can shorten the detection time, save costs, and effectively improve the work efficiency of on-site inspectors. In the development of computer industry, integrated cooling mobile computer cabinet and control system with UPS power assembly, improve the efficiency of computer work, reduce unnecessary power loss, improve the working capacity of computer [2, 3].

According to the power level, UPS power supply can be divided into micro (<3 kVA), small (3 kVA–10 kVA), medium (10 kVA–100 kVA) and large (>100 kVA). At present, UPS is crowded in high-end market in China, and the equipment is expensive. However, UPS minicomputer has been developing slowly for many years, and part of the field and equipment field need miniature UPS power supply, but it has no result.

Therefore, this paper mainly aims at the characteristics of low power, stable output voltage, low harmonic distortion and other requirements, and designs a simple micro single online uninterruptible power supply system.

2 Circuit Design

2.1 System Working Principle

This design uses 220 V AC power supply, lead-acid battery energy storage, power system for flyback power supply.

The rectifier circuit provides stable DC power to the rear inverter module while charging the battery. Moreover, the rectifier circuit includes PFC circuit, which can make more efficient use of energy and provide the power supply to the subsequent circuit stably. After weighing all parties, the inverter circuit chooses EGS002 inverter driving template. The module is small, flexible and high precision. It can invert the input DC current into a sine wave with low distortion rate, and has the functions of over voltage and over current protection, digital display and so on.

The charging circuit is dominated by UC3842 to charging the battery. The switching circuit adopts a two-transistor T-shaped parallel structure, which saves products and is convenient and quick. The lead-acid battery was chosen as the battery. Although the battery is slightly larger, the lead-acid battery has the advantages of large energy storage and high cost performance, which makes it finally adopt the lead-acid battery.

The overall design block diagram is shown in Fig. 1. After the municipal power input, it becomes a relatively stable DC energy through the rectifying circuit, and gives the inverter circuit function while charging the lead-acid battery, Use the unidirectional conductivity of the diode to automatically select the voltage source to achieve the purpose of uninterrupted power supply. The inverter circuit inverts the DC power into 50 Hz sinusoidal AC power. At the same time, the SG3525 chip monitors the line in real time, protects against over voltage and over current, and provides stable and accurate power to the load.

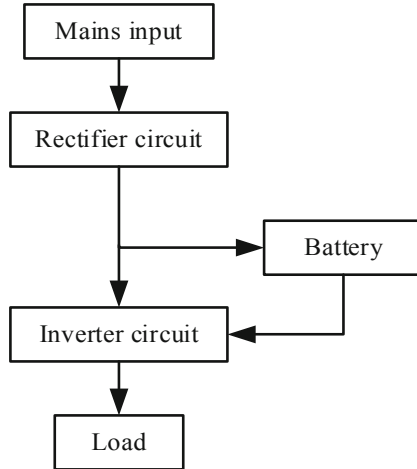


Fig. 1. Diagram of overall design

2.2 Rectifier Circuit

Rectifier circuit, simply is AC into DC circuit.

It has two main functions: First, the Alternating Current (AC) into Direct Current (DC), after filtering supply load, or supply inverter; Second: Provides charging voltage for the battery. Therefore, it also acts as a charger [4].

The rectifier circuit is full-bridge rectifier. DB5010P full-bridge module is used for full-wave rectifier, and stable DC voltage is obtained after second-order filtering. The relevant calculation formula is as follows:

$$U_d = 1.414U_a \quad (1)$$

Where, u_d is the voltage after rectification, and u_a is the effective voltage of alternating current. The AC electric energy outputs half-wave electric energy after passing through the full-bridge rectifier chip; Under the control of UC3842 chip, the switch tube works at high frequency and outputs DC pulse electric energy. The circuit structure adopts the flyback type, the circuit form structure is simple, the energy efficiency is high, and it is suitable for this miniature power supply equipment. Its design block diagram are shown in Fig. 2.

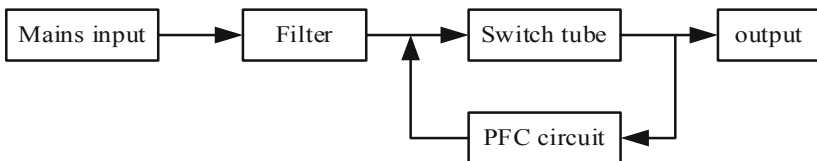


Fig. 2. Diagram of rectifier circuit design

2.3 Switching Circuit

The power supply switching scheme has single-chip control, relay control, static switch and transistor control.

Because the single-chip microcomputer control in programming and wiring is more complex, although it can achieve soft start protection and other functions, but the scheme is not suitable for micro equipment, so it is not adopted. However, relay has a large delay, high price, large volume, is also not the best choice [5].

The static switch is a non-contact switch, it is an AC switch composed of two (SCR) in reverse parallel. The closing and opening of the switch is controlled by the logic controller, it is divided into conversion type and parallel type. Conversion switch is mainly used for two power supply system, its role is to achieve automatic switching from one to the other - road. Parallel switch is mainly used for parallel inverters and mains or multiple inverters [6].

After comprehensive consideration, a structure control similar to a static switch is finally selected, that is, a diode automatic switching circuit, which uses a double transistor to control on and off automatically. The schematic diagram is shown in Fig. 3, where V2 is the dc after rectification, V1 is the battery, and R1 is the load.

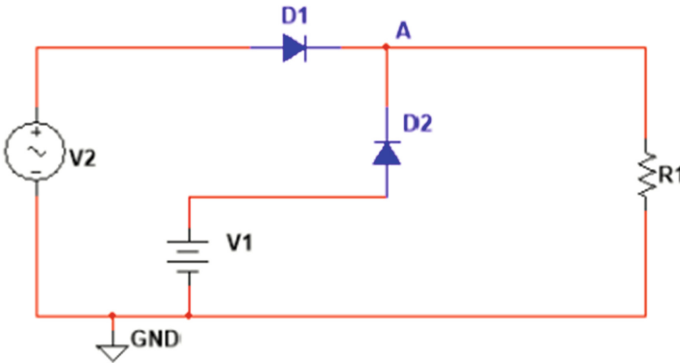


Fig. 3. Switching schematic

When the AC power supply works, D1 turns on, the potential of point A is greater than the voltage of dc power supply, D2 cuts off, and the subsequent electricity is supplied by DC power supply V1. When the dc power supply V1 is cut off, the potential of point A drops. When the node potential is less than the voltage of battery V1, D2 turns on, and the subsequent electricity is supplied by battery V1. At this time, the potential of point A is greater than the potential of DC power supply V2 after the power failure, D1 cuts off, and forms protection for the AC power supply side.

2.4 Inverter Circuit

Combine the IGBT single-phase full-bridge inverter circuit (Fig. 4) to illustrate the PWM modulation method [7].

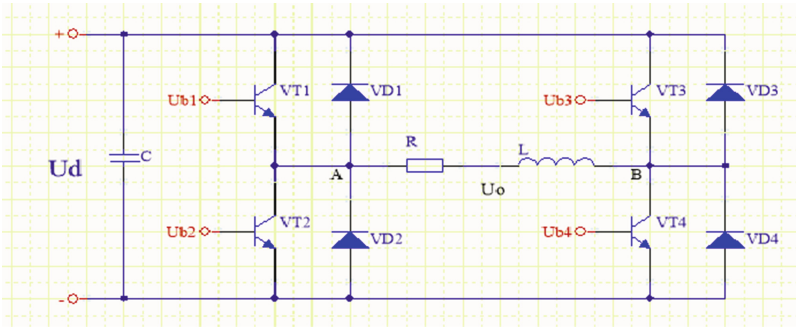


Fig. 4. Full-bridge inverter circuit

V1 and V2 on and off complementary, V3 and V4 on and off complementary. Take positive half cycle of U_0 as an example, V1 is disconnected, V2 is disconnected, V3 and V4 are disconnected alternately. The load current lags behind the voltage. In the positive half of the voltage, the current is positive in part and negative in part. When the load current is positive, V1 and V4 are turned on, U_0 equals U_d . V4 off, load current through V1 and VD3 continued, $U_0 = 0$ load current is negative, V1 and V4 are still on, i_o is negative, in fact, i_o from VD1 and VD4 flow, there is still $U_0 = U_d$. After V4 is turned off and V3 is turned on, i_o continues to flow from V3 and VD1, and $U_0 = 0$. U_0 can get U_d and zero two levels. U_0 negative half cycle, let V2 keep on, V1 keep off, V3 and V4 alternately on and off, U_0 can get $-U_d$ and zero two levels.

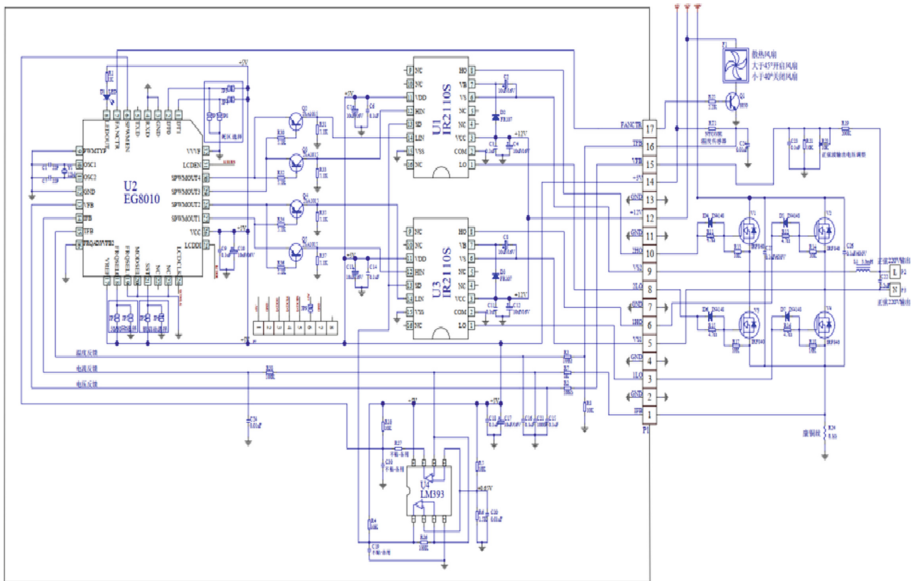


Fig. 5. Driver circuit of EGS002

Refer to the schematic design of EGS002 driver board (Fig. 5) provided in “EGS002 Sine Wave Inverter Driver Board User Manual V1.0” [8]. for the peripheral circuit of EGS002 module.

EGS002 is a pure sine wave inverter driver board, which mainly includes EG8010 and IR2110 chips. EG8010 chip is used to realize the functions of unipolar SPWM modulation output and dead zone, and IR2110 chip is used to realize the functions of driving low power MOS tube [9, 10]. The output sampling is to send the data to EGS002 module after the partial voltage rectification. In case of over voltage or over current, the power can be cut off immediately to protect personnel and line safety.

3 Test Method and Results

During testing, load adjustment rate (SI), voltage adjustment rate (SU) and efficiency (η) will be calculated according to formulas (2), (3) and (4).

$$SI = \frac{|U_0(0.1 \text{ A}) - U_0(1 \text{ A})|}{30} \quad (2)$$

$$SU = \frac{|U_0(43 \text{ V}) - U_0(29 \text{ V})|}{30} \quad (3)$$

$$\eta = \frac{(U_0 \times I_0)}{(U_d \times I_d)} \quad (4)$$

U_0 (0.1a) and U_0 (1 A) are the corresponding output voltage U_0 when the output current I_0 is 0.1a and 0.1a, respectively, during the load adjustment rate test; U_0 (43 V) and U_0 (29 V) refer to the corresponding output voltage U_0 when the input voltage U_1 is 43 V and 29 V in the voltage adjustment rate test respectively.

Measurement Scheme:

- ① Adjust the voltage converter ratio, so that its output voltage is 36 V. Change the load resistance value so that the current changes between 0.1a and 1.0 A, and record the voltage at 0.1a and 1.0 A respectively. The obtained values are recorded in Table 1.
- ② Change the transformer ratio so that the output voltage is 29 V and 43 V respectively. Change the load resistance value so that the current is constant 1 A, measure the output voltage, and calculate the value using the formula and record it in Table 2.
- ③ replace the standby power supply (energy storage device) with 24 V power supply, change the load resistance value to 30 ω , measure the dc power supply current, calculate the η , and record the value in Table 3.

Test results: (average value is taken for 8 times in each group):

Table 1. Test results 1

Project	1	2	3	4	5	6	7	8
$U_0(0.1 \text{ A})$	29.2	31	30.2	29.9	30.6	30.5	30.2	30
$U_0(1.0 \text{ A})$	31	31.2	30.2	30.3	29.2	29	30.5	31.2
S_1	6.0%	0.7%	0.0%	1.3%	4.7%	5.0%	1.0%	4.0%

Average measured: $S_1 = 2.21\%$

Table 2. Test results 2

Project	1	2	3	4	5	6	7	8
$U_0 = 29 \text{ V}$	30.2	30.1	30.3	29.8	30	29.8	29.8	30.2
$U_0 = 43 \text{ V}$	29.9	30	30.2	30.1	30.2	30	29.4	30.1
SU	1.0%	0.3%	0.3%	1.0%	0.7%	0.7%	1.3%	0.3%

Average measured: $S_1 = 0.7\%$

Table 3. Test Results 3

Project	1	2	3	4	5	6	7	8
I_d	1.5	1.25	1.6	1.8	1.3	1.5	1.8	1.3
η	83%	0%	78%	69%	96%	83%	69%	96%

Average measured: $S_1 = 71.75\%$

The following figure can be obtained after the output terminal is connected to the oscilloscope.

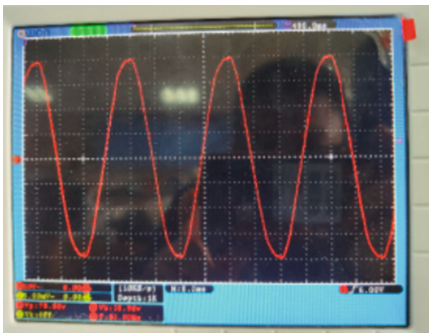


Fig. 6. Sine curve without load

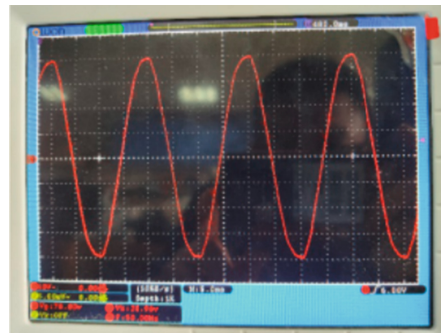


Fig. 7. Sine curve with load

By comparing Fig. 6 and Fig. 7, it can be seen that the adjustment rate of no-load output and on-load output of the power supply reaches the standard.

4 Conclusion

On the premise of reaching the use conditions, this design demonstrates the feasibility of the single online uninterruptible power supply system design with a simple circuit structure, and proves that small UPS equipment can be used. However, EGS002 module is used in the inverter control part of the design, which lacks certain autonomy. In the later improvement, SCM or chip will be used instead to obtain the complete autonomy of the circuit.

Acknowledgements. Supported by Special Project of Education and Teaching Reform Project of Tianjin Agricultural University (2021-A-43) & National Training Program of Innovation and Entrepreneurship for Undergraduates (202110061005) & Plan Project of Hebei Innovation Capability Improvement (199A1912H).

References

1. Yi, X.: Design of single-phase online uninterruptible power supply based on STM32. *J. Phys.* **1802**(3), 591–597 (2021)
2. Qiufeng, L., et al.: Design and application of portable fault indicator detector based on internet of things. *IOP Conf. Ser. Mater. Sci. Eng.* **486**(1), 14–24 (2019)
3. Dong, Z.: Principle and design of single-phase online uninterruptible power supply. *Electron. Technol. Softw. Eng.* **2021**(10), 211–213 (2021)
4. Hassaine, L., Olias, E., Quintero, L, et al.: Overview of power inverter topologies and control structures for grid connected photovoltaic systems. *Renew. Sustain. Energy Rev.* **30**, 796–807 (2014)
5. Thavu, B.: Micro controller based current fed dual bridge DC-DC converter. *Int. J. Emerg. Trends Electr. Electron.* **1**(4), 4–31 (2013)
6. Eckel, H.G., Fleisch, K.: Turn-off behaviour of high voltage NPT- and FS-IGBT. In: 13th International Power Electronics and Motion Control Conference (IPEMC), pp. 100–102 (2017)
7. EGS002 Sine Wave Inverse Drive Board User Manual V1.0
8. Choi, U.M., Lee, H.H., Lee, K.B.: Simple neutral-point voltage control for three-level inverters using a discontinuous pulse width modulation. *Trans. Energy Convers.* **28**(2), 434–443 (2013)
9. Qi, Y., Li, P., Huang, Z.Y., et al.: Feedforward control of output current for three phase voltage source inverter (VSI) with transformer. In: Power Electronics Conference and Exposition (APEC), pp. 2318–2322 (2014)
10. Oliveira, T., et al.: Finite control set model predictive control for paralleled uninterruptible power supplies. *Energies* **13**(13) (2020)



Fault and Detection of Electrical Control System Based on Intelligent Algorithm

Ling Liu^(✉)

Wuhan Railway Vocational Collage of Technology, Wuhan 430223, Hubei, China

Abstract. The electric control system (Electric Control System, ECS) has developed into the core of the automation control of electrical equipment, and it is also a necessary system for the control of modern electrical equipment. The advancement, reliability, and sensitivity of ECS have become an important guarantee for the efficient, stable and safe operation of the power industry. Therefore, it is very important to improve the ability to control the operation status of ECS and the ability to detect ECS failures. Through research, this paper uses intelligent algorithms to realize the fault detection of electrical control system and improve the fault diagnosis level of BP neural network. This article analyzes the source of ECS failure in detail. Based on intelligent algorithms, this paper proposes an ECS fault diagnosis model and fault signal detection method based on BP neural network. By comparing the experimental data with the reference value, the accuracy rates of the wavelet change method, graph theory method, hybrid leapfrog algorithm and wolf pack algorithm are 76.8%, 79.6%, 84.2%, and 86.9%, respectively. However, the fault of the algorithm proposed in this paper the detection accuracy rate is 94.1%. This proves the feasibility of the algorithm and the accuracy of the calculation results in this paper, and achieves the goal of ECS fault detection.

Keywords: Fault diagnosis · Neural network · Diagnosis efficiency · Electrical control system

1 Introduction

With the continuous upgrading of electronic information software and hardware, ECS has also developed and improved in practice [1, 2]. In the past, low-voltage, small-capacity electrical equipment has gradually developed into the current high-voltage, high-parameter, large-capacity, and high-efficiency electrical unit unit. The electrical operation control is transformed from manual to automatic and intelligent control [3, 4]. During the development of ECS, the system has made great progress in design, installation, commissioning, operation, maintenance, and overhaul [5]. However, once the ECS has a problem, it will directly affect the safe and stable operation of related equipment, so the study of ECS fault detection is imminent [6].

Regarding the research of ECS or fault detection, many scholars have conducted multi-angle investigations. For example, Chen CY designed a distributed power remote control system for building power and green power generation [7]; Yasser is based on

space vector, studied the method of electrical fault detection of the motor during the speed change [8]; Kaleybar HJ studied the electrified railway grid controller and proposed an intelligent control method [9]. It can be seen that, based on intelligent algorithms, this paper proposes an ECS fault diagnosis method based on BP neural network, which has practical significance.

This article first conducts an in-depth analysis and sorting of the sources of ECS failures. Then, based on intelligent algorithms, combined with electrical fault information, an ECS fault diagnosis model and fault detection method based on BP neural network is proposed. Finally, through experiments, the feasibility of the algorithm in this paper and the accuracy of the calculation results are verified.

2 ECS Fault and Detection Based on Intelligent Algorithm

2.1 Source of ECS Failure

This article divides the ECS's fault sources into four parts: power grid parameters, hardware equipment, communication media, and drivers. The traditional fault monitoring and diagnosis system only analyzes and processes power grid parameter faults.

(1) Grid parameters

For power supply companies, the impact of grid parameter failures on ECS is the most direct. Once a failure occurs, it will directly affect the safety and reliability of the power supply system. For power users, power parameter failure will directly affect the normal operation of the equipment, and even cause the consequences of equipment damage. Therefore, power grid parameters are important monitoring and diagnosis objects in terms of fault diagnosis. Grid parameter faults can be subdivided into two types: the first is basic number faults, such as digital input faults, relay control faults, voltage deviations, current deviations, frequency deviations, power deviations, power factor deviations and electrical energy deviations, etc.; the other is a power quality parameter. Therefore, new energy sources such as wind and hydropower are connected to the grid, and rectification and filtering technologies benefit mankind. At the same time, they will also cause great pollution to the grid environment. Therefore, the detection of current and voltage waveforms, harmonic interference, and current and voltage unbalance and other faults related to power quality parameters are very important.

(2) Hardware equipment

There are a wide variety of hardware equipment integrated in digital intelligent substations, such as intelligent electronic equipment such as DC, EPS (emergency power supply), transformers, thermostats, comprehensive protection devices, and smart power meters. If these devices are damaged, power-off, and the device model does not match the driver, it will affect the detection of grid parameter failures and the upload of failure information. Therefore, the importance of the fault information of the hardware device itself is becoming increasingly prominent.

(3) Communication medium

Only after the upper computer in the ECS communicates with the lower computer equipment can the grid parameter fault information and hardware equipment fault information be obtained. Once the communication medium fails and the communication between the upper computer and the lower computer is interrupted, the fault information cannot be obtained correctly and in real time. The failure of the communication medium can be specifically divided into the interruption of the communication medium, the communication medium being interfered, the communication port failure of the upper computer or the lower computer equipment, and the manual connection error [10].

(4) Driver

In ECS, the fault monitoring and diagnosis system is inseparable. Therefore, the software program that is most likely to fail is the driver of the hardware device. The driver obtains the fault information transmitted by the hardware device according to the communication protocol. If the driver is stuck, memory leaks, abnormal thread count, and abnormal I/O read/write, it will cause the user to misjudge the fault, or even think that no fault has occurred. Therefore, it is very necessary to monitor and diagnose the fault of the driver. In particular, the driver memory leak problem will consume a lot of memory resources after the system runs for a long time, which is an important reason for the driver to get stuck.

2.2 BP Neural Network Fault Diagnosis Method

When the ECS fails to operate normally, maintenance personnel need to use relevant monitoring instruments to monitor the operating parameters of the system, and determine the cause of the system's failure to operate normally based on the analysis of the operating parameters, but the detection method is slow and inefficient, and it is difficult to achieve Real-time detection of equipment failures. This article will use the BP neural network algorithm to analyze ECS faults, so as to quickly troubleshoot ECS faults.

(1) Extraction of fault information

Through the preliminary demodulation of the vibration signal, analysis of the system's vibration signal and three-phase current signal spectrum low-frequency region, it is found that the energy of the signal is mainly distributed in the low-frequency region; the characteristics of the four signal channels are extracted from each spectrum. Since the extracted features change with the change of the motor speed, the rotation frequency is used as the reference for frequency region division; respectively use $S[n]$, $IA[n]$, $IB[n]$, $IC[n]$, $n = 1, 2, \dots, N$ represents the discrete form of vibration and three current signals, where N is the sampling point of each frame of signal; the frequency spectrum of these signals are respectively denoted as $PS[n]$, $PIA[n]$, $PIB[n]$, $PIC[n]$, $n = 1, 2, \dots, N$; frequency resolution calculation $\Delta f = f_s/N$. For each fault signal, 10 features are extracted from the frequency spectrum, and a total of 40 features are obtained, as shown in formulas (1) and (2):

$$\begin{cases} F[j] = \frac{1}{K_2-K_1} \sum_{i=K_1}^{k_2} PS[i], F[10 + J] = \frac{1}{K_2-K_1} \sum_{i=K_1}^{k_2} PIA[i] \\ F[20 + J] = \frac{1}{K_2-K_1} \sum_{i=K_1}^{k_2} PIB[i], F[30 + J] = \frac{1}{K_2-K_1} \sum_{i=K_1}^{k_2} PIC[i], \end{cases} \quad (1)$$

$$K_1 = \text{round}[\frac{(j - 0.5)f_r}{\Delta f}], K_2 = \text{round}[\frac{(j + 0.5)f_r}{\Delta f}], j = 1, 2, \dots, 10 \quad (2)$$

In formulas (1) and (2), 1 is the rotation frequency estimated by the Hall sensor pre-installed on the brushless DC motor. For deterministic motor conditions, due to the uncertainty of the signal and the change of frequency, the characteristic value of each signal group changes within a certain range. However, since the frequency spectrum distribution is mainly determined by the state of the motor in the ECS, the extracted features can reflect the operating state of the system. When ECS has different electrical or mechanical failures, these characteristic values will vary with the type of failure. In addition, in order to eliminate the influence of signal amplitude, the vibration signal and three current signals in each motor state are normalized, as shown in formula (3):

$$\begin{cases} \text{Vibration signal: } F[j] = f[j]/\max(F[j]), j \in [1, 10] \\ \text{Vibration signals: } F[j] = f[j]/\max(F[j]), j \in [11, 40] \end{cases} \quad (3)$$

(2) BP neural network fault diagnosis model

Assuming that n, m, and q are the number of neurons in the input layer, hidden layer, and output layer of the BP neural network, v_{ij} is the weight from the input layer to the hidden layer, b_1 is the threshold of the hidden layer, and w_{jk} is the hidden layer. The weight from layer to output layer, b_2 is the threshold of the output layer, $h()$ and $f()$ are the transfer functions used by the hidden layer and output layer nodes respectively, $X = (x_1, \dots, x_i, \dots, x_n)^T$ is the input of the network, and $O = (o_1, \dots, o_k, \dots, o_q)^T$ is the output of the network, then The input and output relationship of the neural network can be as shown in formula (4):

$$O_k = f[\sum_{j=1}^m w_{jk}h(i = 1 \sum_{i=1}^n v_{ij}x_i - b_1) - b_2 \quad (4)$$

In formula (4), $j = 1, 2, \dots, m$; $k = 1, 2, \dots, q$; the total error function output by the network is shown in formula (5):

$$E = \frac{1}{p} \sum_{p=1}^p \sum_{k=1}^q (d_k^p - O_k^p)^2 \quad (5)$$

In formula (5), p represents the total number of samples contained; d represents the output value that people expect from the network. The stator current signal of the motor under different fault states, the current signal contains the information of the fault, and the fault feature is extracted by means of wavelet packet. The fault feature vector is used as the input of the neural network model, and the output of the fault diagnosis model is the fault type.

3 Fault Diagnosis Test

3.1 ECS Fault Diagnosis

In order to verify the feasibility and effectiveness of the ECS fault diagnosis method based on BP neural network, a fault diagnosis simulation test was carried out on the MATLAB platform for the simultaneous faults of multiple hardware devices.

At present, deep learning and neural network frameworks represented by TensorFlow are gradually being applied to machine learning research and become the mainstream. This article also chooses to build a special ECS fault feature extraction model in the TensorFlow framework. In practical applications, if the memory environment is tight, this model can be easily converted to C++ or R language. In addition, Python and C++ will not have interface problems with industrial reality, and are more in line with industrial automation standards.

The process of implementing neural network fault diagnosis model training can be accelerated by GPU, so TensorFlow can choose the CPU version or GPU version, but for the GPU version of TensorFlow, you need to use NVIDIA's graphics card and the supporting GPU acceleration software CUDA.

3.2 Experimental Data

After preprocessing the fault data obtained by ECS, the interference of other factors is eliminated, and the abnormal heating fault of the ECS thermostat is diagnosed.

All data files are in Matlab (*.mat) format. Digital data collects 15,000 samples per second. Use torque sensors/encoders to collect speed and power data and record them manually.

4 Comparison of Experimental Results of Each Method

4.1 Fault Detection Accuracy

Traditional ECS fault diagnosis methods include wavelet transform method, K-means method, hybrid leapfrog algorithm and wolf pack algorithm. In order to verify the effectiveness and superiority of the algorithm proposed in this paper, the above four methods are used for ECS fault detection. The experimental results of each method are shown in Table 1.

Observing Fig. 1, it is found that the accuracy rates of wavelet change method, graph theory method, hybrid leapfrog algorithm and wolf pack algorithm are 76.8%, 79.6%, 84.2%, and 86.9%, respectively, and the fault detection accuracy of the algorithm proposed in this paper is 94.1%. This shows that the BP neural network fault detection method proposed in this paper can greatly improve the accuracy of ECS fault diagnosis.

Table 1. Failure detection results of each method (%)

Experimental method	Accuracy	Recall rate
Wavelet transform	76.8	83.1
Graph theory	79.6	87.6
Hybrid Leapfrog Algorithm	84.2	89.4
Wolf pack algorithm	86.9	90.1
This paper proposes an algorithm	94.1	96.3

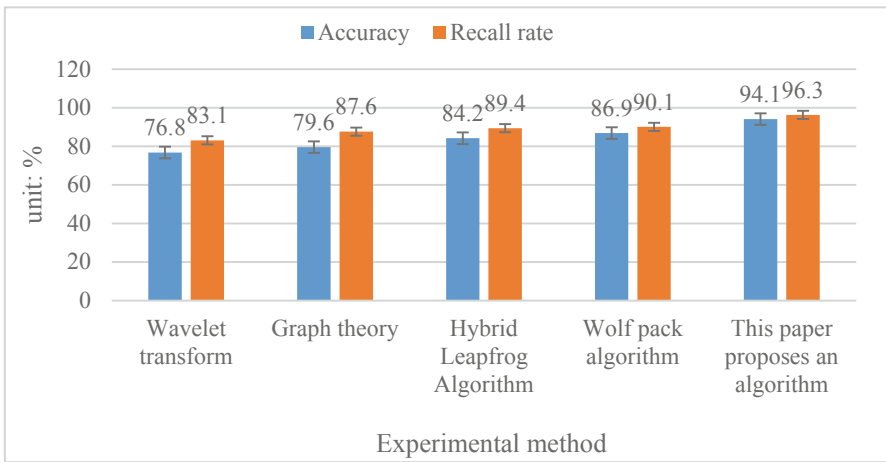


Fig. 1. Failure detection results of each method

4.2 Different Algorithm Fault Detection Degree and Detection Time

The degree of failure refers to the proportion of the area of the failure area in the total area of the target. The algorithm calculates the percentage of the target area of the fault area where the thermostat has abnormal heating, which is helpful to further determine the degree of the fault and help the inspector make more accurate judgments and treatment measures. The failure degree and iteration time results of each algorithm are shown in Table 2: The failure degree of the actual sample data is 17.6%, and the failure degree calculated by the wavelet change method, graph theory method, hybrid leapfrog algorithm and wolf pack algorithm are respectively 30.2%, 29.3%, 41.1%, 23.6%, while the degree of failure calculated by the algorithm proposed in this paper is 16.9%.

It can be seen from Fig. 2 that the difference between the experimental results of this paper using the improved algorithm and the reference value is 0.7%, and the calculation error in the above several algorithms is the smallest. And the fault detection of the sample data only takes 4.13 s, the time is the shortest, and the efficiency is the fastest. This shows that the accuracy and efficiency of the algorithm proposed in this paper can meet the requirements of practical applications.

Table 2. Different algorithm fault detection degree and detection time

Experimental method	Degree of failure (%)	Time (s)
Wavelet transform	30.2	4.57
Graph theory	29.3	4.29
Hybrid Leapfrog Algorithm	41.1	4.65
Wolf pack algorithm	23.6	4.76
This paper proposes an algorithm	16.9	4.13
Reference	17.6	–

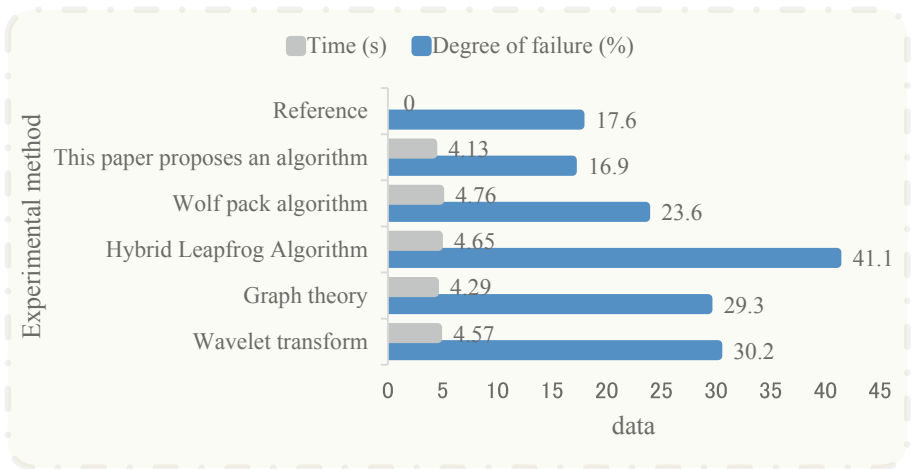


Fig. 2. Different algorithm fault detection degree and detection time

5 Conclusions

ECS means that after collecting on-site electrical information, sharing information through the communication network, monitoring electrical equipment at the operator station, and issuing control instructions to operate the electrical equipment through the corresponding actuator. Once the ECS fails, it will seriously affect the operation of electrical equipment. This article first conducts an in-depth analysis of the ECS operating requirements and fault sources, and then proposes an ECS fault detection method based on BP neural network, which verifies the good performance of the algorithm from the results.

References

1. Swain, K.P., De, M.: A novel electrical proximity index for voltage control in smart distribution system. *Electr. Power Syst. Res.* **172**, 50–62 (2019)

2. Love, P., Zhou, J., Matthews, J.: Project controls for electrical, instrumentation and control systems: enabling role of digital system information modelling. *Autom. Constr.* **103**, 202–212 (2019)
3. Spagnolo, C., Sumsurooah, S., Hill, C.I., et al.: Finite state machine control for aircraft electrical distribution system. *J. Eng.* **2018**(13), 506–511 (2018)
4. Gaspar, J.E., Kamarlouei, M., Sinha, A., et al.: Analysis of electrical drive speed control limitations of a power take-off system for wave energy converters. *Renew. Energy* **113**, 335–346 (2017)
5. Dansoko, M., Nkwawo, H., Floret, F., et al.: Marine turbine system directly connected to electrical grid: experimental implementations using a nonlinear and robust control. *Ocean Eng.* **149**, 260–267 (2018)
6. Yasser, G., Rossi, C., et al.: A diagnostic space vector-based index for rotor electrical fault detection in wound-rotor induction machines under speed transient. *IEEE Trans. Ind. Electron.* **64**(5), 3892–3902 (2017)
7. Kaleybar, H.J., Kojabadi, H.M., Fazel, S.S., et al.: An intelligent control method for capacity reduction of power flow controller in electrical railway grids. *Electr. Power Syst. Res.* **165**, 157–166 (2018)
8. Amber, K.P., Ahmad, R., Aslam, M.W., et al.: Intelligent techniques for forecasting electricity consumption of buildings. *Energy* **157**, 886–893 (2018)
9. Fan, S., Liu, Z., Feng, G., et al.: Dynamic process analysis and voltage stabilization control of multi-load wireless power supply system. *Energies* **14**(5), 1466 (2021)
10. Quach, N.T., Chae, S.H., Song, S.H., et al.: Frequency and voltage control strategies of the Jeju island power system based on MMC-HVDC systems. *J. Power Electron.* **18**(1), 204–211 (2018)



Design and Implementation of Graduate Student Enrollment Management Information System Based on Rsync Algorithm

Cui Liu^(✉)

Changchun Institute of Technology, Changchun 130021, Jilin, China

Abstract. With the rapid development of computer technology, information technology and networks have become important functions of our time. Various social tasks increasingly rely on computer networks to achieve easy, fast, and standardized management goals. As one of the key challenges that the school faces every year, the admission of graduate students is not only an effective way to improve the quality of admission, but also provides a wide range of questions and exchanges for candidates through the provision of information, which is also an important role. By providing candidates with a full range of services, it helps to effectively promote the role of publicity and enhance the school's external image. The purpose of this paper is to study the design and implementation of a graduate student enrollment management information system based on the Rsync algorithm. Based on the analysis of the Rsync algorithm process, database tools and options, and system non-functional requirements, the Rsync algorithm-based graduate student enrollment management information system is carried out. Design and test the performance of the system. The test results show that the system designed in this paper has stable performance and supports no less than 300 concurrent users.

Keywords: Rsync algorithm · Graduate enrollment · Management information system · Enrollment process

1 Introduction

The rapid development of information and network technology has brought major changes to the work of all walks of life in society. Office work is more convenient and quicker, reducing work error rate and greatly improving work efficiency. In order to improve admission efficiency, reduce workplace errors, reduce admission costs, and strictly control the admission process, the University Graduate Admissions Office outlined the requirements for establishing a graduate admissions management information system [1, 2].

With the continuous improvement of school conditions, the number of students enrolled in the school continues to increase, and the information brought by the registration process to the exam also increases. Some scholars pointed out that as the amount of information continues to increase, it is necessary to use the office information system

to implement enrollment information management and improve the efficiency of enrollment work. In addition, the competition among universities for quality undergraduates has become increasingly fierce. However, the admissions unit does not have effective means to obtain information for high-quality students [3, 4]. Especially in the process of adapting to postgraduate exams, due to the information asymmetry between candidates and colleges, there are many uncertainties in the re-examination work. The information system can effectively implement standard information management mode, statistical scientific data technology and effective data retrieval mechanism, and reduce the burden on registered personnel. Some scholars pointed out that “enrollment informationization” is an innovative practice of “Internet information management + enrollment” that combines Internet technology with college registration management [5, 6]. From a technical point of view, some researchers believe that enrollment informatization has two main characteristics: digitalization and networking. Digitization transforms educational information into accurate digital information, and performs advanced information classification management. Networking can realize the sharing of information resources. Using the Internet as a means to reduce the time and space constraints of sending and sharing information, making it more convenient for employees to share. At the same time, eliminate the negative impact of information asymmetry on candidates and enrollment units [7, 8]. There are also some researchers’ traditional campus information management methods involving information transfer between different departments and institutions, mainly through face-to-face communication, telephone voice communication, paper document submission, etc. The information processing efficiency is low, and the information that is difficult to understand is effectively classified. And management, the integrity and reliability of the data cannot be guaranteed when transmitting information, and there are certain risks, so the cost of information management will increase significantly [9, 10]. By introducing information management methods, it is possible to solve the efficiency problems caused by information transmission media and avoid the weaknesses of traditional working methods.

On the basis of consulting a large number of relevant references, this paper designs a postgraduate enrollment management information system based on the Rsync algorithm process, database tools and options, and system non-functional requirements. The system mainly includes four modules, namely the registration management module and the enrollment management module, retest management module and admission management module. Finally, the performance of the system is tested to verify whether the system meets the requirements of this article.

2 Design and Implementation of Graduate Student Enrollment Management Information System Based on Rsync Algorithm

2.1 Rsync Algorithm Flow

The Rsync algorithm is currently the most mainstream remote file synchronization algorithm. The file synchronization rate is also quite fast, because we use incremental synchronization [11, 12]. The Rsync synchronization algorithm can count the number of messages that need to be sent in a short time. The general process of Rsync synchronization calculation includes:

Suppose that the same files A and B are synchronized between two computer systems u and v, computer system u accesses file A, and computer system two accesses file B. Now, we want to synchronize file B and file A, and assume that the network bandwidth between host u and v is too small. Then use the rsync algorithm to complete the following five processes:

- (1) v Divide file B into a set of unique data blocks, the size of which is specified in L bytes, and the last block may be smaller than L;
- (2) vhost calculates two control sums for each divided data block. One is the total number of weak 32-bit control cycles, and the other is the total number of strong 128-bit MD4 controls;
- (3) v sends these verification results to u;
- (4) Host u calculates a 32-bit cyclic checksum of L blocks from the beginning of file A, and checks whether it is the same as the weak checksum of file B. If it is the same as the OMD4 code, the data block exists in B, and the file index of A is moved by L bytes. Otherwise, move 1 byte. This task can be completed quickly by using the characteristics of the loop detection algorithm.
- (5) Host u sends a series of commands to host v to create a backup B' of file A in u. Each command here either proves that a specific A data block already exists in file B without resending (send the command shortly after the strong checksum is the same), or in file A. It is a data block that exists but is not in file B.

The 32-bit loop control used in the algorithm and its inspiration comes from Mark Adler's adler-32 control total algorithm. The calculation formula includes formula (1), formula (2) and formula (3):

$$a(k, l) = \left(\sum_{i=k}^l X_i \right) \bmod M \quad (1)$$

$$b(k, l) = \left(\sum_{i=k}^l (l - i + 1)X_i \right) \bmod M \quad (2)$$

$$s(k, l) = a(k, l) + 2^{16}b(k, l) \quad (3)$$

Where $s(k,l)$ is the checksum of bytes from X_k to X_l .

2.2 Database Tools and Options

At present, the main database development tools mainly include Access database, SQL Server database, and Oracle database.

- (1) Microsoft Access is one of the main components of the automated management software in the Office system. A small database system. It has the advantages of large resolution mode, strong versatility, and simple use. But the disadvantage is that it is a single-user operating system, which may be accessed by many applications, so it can only be used on the Windows platform.

- (2) SQL Server is currently the most popular relational database. It not only inherits the powerful functions of the BackOffice series of products, but also extends it appropriately. It is welcomed by the general public for its graphical user interface, rich programming interface tools, strong scalability, and support for large-scale websites and enterprise information processing systems. In addition, the client/server architecture divides the application into multiple tasks, which can be run as separate programs.
- (3) Oracle Database is a database product released by Oracle, with strong compatibility, portability, high connectivity and flexibility.

In contrast, the SQL Server database not only meets the needs of the graduate enrollment management information system, but also has low development costs, short cycles and easy maintenance. Taking the graduate school as an example, the total number of applications for various graduate students can reach about 50,000 each year. Therefore, the graduate school management information system is not a management information system that requires a very large database. Therefore, this system uses SQLServer database for development.

2.3 System Non-functional Requirements

(1) Security

Important information such as student personal information and control information needs to be kept confidential for most users, so security requirements are high. CSRF (intermediate request), XSS (intersite scripting attacks), batch injection of SQL Actions, DDos attacks and other issues need to be resolved.

(2) Reliability

The system contains a large number of users. According to the registration data of the past few years, the number of users is estimated to be around 10,000. The peak visits of student pages (such as registration pages) can reach hundreds of students. On the review page, you need to query dozens to thousands of student information at the same time, so you need to choose a web server suitable for multi-threading.

(3) Fault tolerance

The users of the system include not only students, but also teachers of all ages. Therefore, the system also needs a certain degree of fault tolerance, and can be rolled back on some non-critical steps, or the same rollback effect can be achieved through certain operations.

(4) Scalability

The service object of this system is the registration unit of the master degree, and the registration difference of each unit is required. The system should also have an interface that adapts to the situation of different registration units, and proposes secondary development. A set of process recommendations and some flexible configuration system settings.

(5) Ease of use

As mentioned earlier, simplicity and ease of use is one of the non-functional requirements of the system, because the system has a large user population, with many students and users. Therefore, the system needs to provide a humanized interactive interface and more

complete prompts and guidance information. At the same time, error feedback information is removed, and feedback channels are provided for administrators to facilitate user interaction and solve problems in a timely manner.

3 Experiment

3.1 Registration Management Module

Candidates fill in detailed information about candidates based on this year's registration list, which mainly includes basic information, degree and academic information, registration information and contact information, to provide a reference for secondary colleges to monitor the number of candidates. Due to the uncertainty of object-oriented records, the recorded data often contains missing or incomplete information. This information is not available to the admissions staff. The system only needs to save the complete information of the candidates, so the system needs to filter this part and collect it. After completing the candidate data, you need to activate the application and create a registration number, which will later be used as candidates for on-site verification. At the same time, only the candidate who created the registration number can accept the information review by the admissions office and the college secretary. Then manage the key points of completing the information, including important links such as undergraduate colleges, transfer students, transfer professional score limits, and student registration numbers.

3.2 Admissions Management Module

The enrollment management module is mainly used to collect candidate information and publish enrollment information. This module will be managed by the Admissions Office of the Graduate School, and will be completed in conjunction with the secondary schools and departments and all candidates. The enrollment management part includes three parts: the establishment of the professional enrollment list, the recommendation and acceptance of excluded candidates, and the management of candidate registration. The Admissions Office will announce and compile the enrollment list for the year before enrolling graduate students. The enrollment list is mainly designed for relevant examination information, such as enrollment majors, number of students to be enrolled, examination subjects, etc. The recommendation and acceptance of non-promotional students mainly includes the self-recommendation of non-promotional students, the review by teachers of colleges and departments, and the acceptance of school opinions, etc., to generate an information database of non-promotional students. Candidate registration management is mainly to collect candidate information and create database information such as candidate registration database and picture library.

3.3 Retest Management Module

The head of the Research and Admissions Office determined the school's ranking line based on the subject category and the number of students enrolled in this year's exam. The department reviews the list of candidates above the school score, and determines

the list of candidates who can participate in the review based on the number of students who can be admitted to their majors. At this stage, transfer students from other faculties and schools can be accepted. Candidates with special circumstances will be highlighted in the review list. The location of the comprehensive written examination is determined by the Research Admissions Office, the time and location of the re-examination subject examination are uniformly determined by the department, and the re-examination venue is the responsibility of the department. Candidates can download the re-examination notice and physical examination form online after arranging the re-examination. Each department then organizes the candidates to retest under their own department. After the re-examination is completed, the department secretary enters the re-examination results of the department candidates and selects the final weighted score. The secretary of the department will determine the type of admission based on the student's final grades, score admission, and the Office of Research and Admissions will review the admission list. Students who have not been admitted can adjust. The process is shown in Fig. 1:

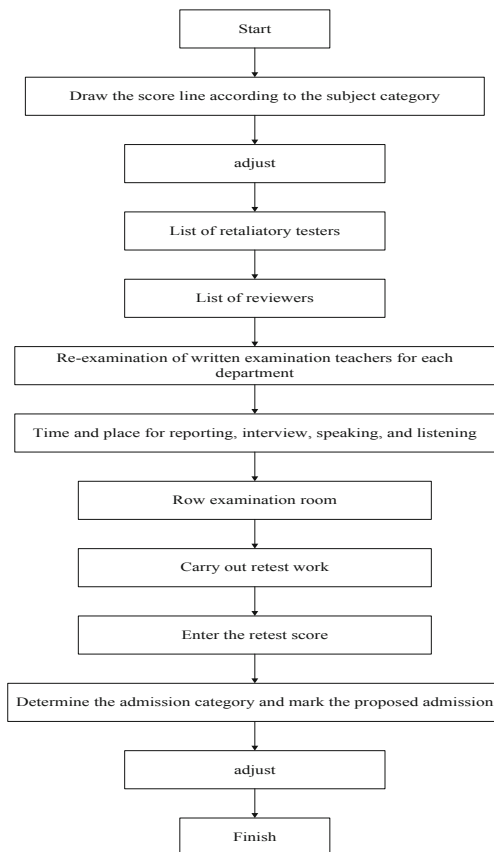


Fig. 1. Flow chart of retest management module

3.4 Admission Management Module

The basic procedures and operations of the admission management module for master enrollment and exemption admission are basically the same. The education unit enters the candidate admission information through the preliminary admission management, and indicates and stamps it, creates the student number, and after the approval is completed, the official reply is submitted to the Ministry of Education, but the master's admission management is one step beyond the national admissions inspection.

The function of saving the input information. First enter the candidate's approval related information, then send and receive the entered related information through the controller, assign the data to the `Examinee_enroll_info` object, store the data in the database through logic and data levels, and get the result.

Submit input data and create student ID. Candidates who have passed the admission assessment can prepare student numbers for those who have passed the admission assessment, and compile information such as the examination method and professional code of the candidate admission information that requires candidate admission information. Therefore, you need to obtain this information before numbering.

4 Discussion

System performance testing is mainly to evaluate the performance of the system under extreme environments. In this way, it is obvious whether the developed system meets the required performance goals. In the test, multiple users are simulated to access the same view page through LoadRunner Controller, and performance indicators are recorded. The results are shown in Table 1 and Fig. 2.

Table 1. Performance test results

	Registration management module	Admissions management module	Retest management module	Admission management module
50	0.243	0.347	0.351	0.210
100	0.574	0.655	0.633	0.457
150	0.861	0.952	0.826	0.791
200	0.958	1.262	1.314	0.964
250	1.651	1.533	1.853	1.543
300	1.985	1.864	2.071	1.798

It can be seen from the test results in Table 1 and Fig. 2 that the system is stable and supports no less than 300 concurrent users. When 300 people were tested concurrently, the test passed.

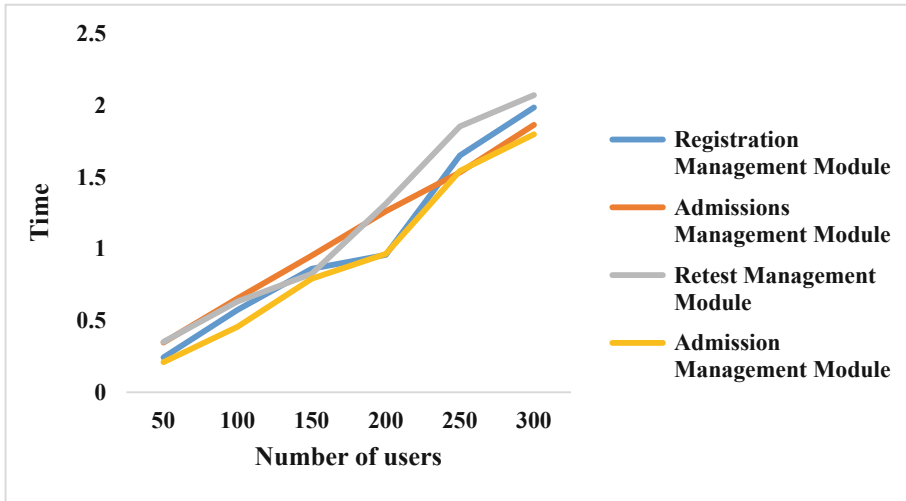


Fig. 2. Performance test results

5 Conclusions

The postgraduate admissions management system is not only convenient for admissions officers, but also for students to apply for graduate school entrance exams. The postgraduate enrollment management system improves the efficiency of the school's enrollment work, reduces the manpower and material resources of the school's enrollment, and allows the school to promote postgraduate enrollment during the annual enrollment period.

References

1. Katie, J.: Grit vs. grades: Veterinary schools face potential flaws in admissions system. *DVM360*, 49(6), 1,10,12–14 (2018)
2. Reid, A.A.: Measuring correctional admissions of aboriginal offenders in Canada: a relative inter-jurisdictional analysis. *J. Law Soc. Pol.* **31**(1), 3 (2019)
3. Noronha, S., Sangli, S., Hoskote, S.: A scoring system to decide ICU admissions in patients with substance overdose. *Chest* **154**(4), 312A (2018)
4. Hong, Y.L., Lee, J., Lee, S.M., et al.: Effect of a rapid response system on code rates and in-hospital mortality in medical wards. *Acute Crit. Care* **34**(4), 246–254 (2019)
5. Zainudin, F., Ismail, K.: Measuring Malaysia School Resource Centers' standards through iQ-PSS: an online management information system. *IASL Ann. Conf. Proc.* **2010**(1), 20 (2021)
6. Ssg, A., Nb, B., Ey, B., et al.: The impact of an integrated electronic immunization registry and logistics management information system (EIR-eLMIS) on vaccine availability in three regions in Tanzania: a pre-post and time-series analysis. *Vaccine* **38**(3), 562–569 (2020)
7. Hafalir, I.E., Hakimov, R., Kübler, D., Kurino, M.: College admissions with entrance exams: centralized versus decentralized. *J. Econ. Theory* **176**, 886–934 (2018)

8. Roach, A., Rose, A., Be Iers-Jones, K., et al.: Incorporating group interviews into holistic review in baccalaureate nursing school admissions. *Nurs. Educ. Perspect.* **40**(2), 1 (2018)
9. Scott-Railton, T.: Shifting the scope: how taking school demographics into account in college admissions could reduce K-12 segregation nationwide. *Yale Law Policy Rev.* **36**(1), 5 (2018)
10. Fischer, P.R., Bostwick, J.M.: Intentionality in medical school admissions in the COVID-19 era. *Mayo Clin. Proc.* **95**(11), 2306–2308 (2020)
11. Quinn, C., Leon, M.D., Clinchot, D.M.: Strategies for achieving diversity through medical school admissions. *J. Health Care Poor Underserved* **29**(1), 9–18 (2018)
12. Mian, O., Hogenbirk, J.C., Marsh, D.C., et al.: Tracking indigenous applicants through the admissions process of a socially accountable medical school. *Acad. Med.* **94**(8), 1211–1219 (2019)



Information System of Constructive English Teaching Platform Based on RBF Algorithm

Jing Gao^(✉)

Shandong Transport Vocational College, Weifang, Shandong, China

Abstract. With the rapid development of information technology, network technology has become an indispensable part of human survival and progress. In teaching, the use of computer-aided teaching system can well solve the problems of large amount of information and limited knowledge for teachers and students under the traditional education mode. The information construction of English curriculum based on RBF algorithm is an important and innovative work. Firstly, this paper first introduces the objectives of constructive English teaching and the connotation of learning view of constructivism theory, then introduces the characteristics of English teaching informatization, and designs and develops the information system of constructive English teaching platform according to RBF algorithm. Finally, the running performance of the system is tested. The test results show that the maximum response time and minimum response time of the system are about 1 ms. The maximum response time and minimum response time when processing sessions are maintained at about 1–2 ms. From the overall point of view of the system, this platform realizes the management of student information, course courseware and teacher users, and can effectively interact data in the network environment.

Keywords: RBF algorithm · Constructive English · Teaching platform · Information teaching

1 Introduction

With the deepening of China's basic education reform, the teaching mode and content of schools have changed greatly. In information construction, information technology has become a new and scientific technology. It is one of the main tasks for students to choose and acquire information resources as learning subjects. Therefore, it is of great significance to construct an efficient classroom platform [1, 2].

Many scholars at home and abroad have conducted relevant research on the informatization of constructive English teaching platform. In foreign countries, some countries have combined information technology with other disciplines for integrated development and achieved some results. In China, the research on the informatization of teaching platform mainly focuses on network technology and computer application. Among them, network technology includes multimedia, virtual reality and other methods [3, 4]. However, China makes an exploratory research on the construction of educational

resources by learning from foreign advanced experience. In the information technology and subject curriculum standards, some scholars believe that education reform should be student-centered and teacher led [1, 5]. The domestic research on the informatization construction of English teaching platform mainly focuses on the following aspects: first, it puts forward some requirements for teachers, students and school managers, such as building an information sharing center. The second is to build a distance education environment based on the combination of multimedia network technology and computer hardware resources. The third is to establish a set of system functional modules with strong pertinence and high operability to meet the learning needs of different types of college students. The fourth is the software program developed according to the information design principle of English teaching platform, including data memory, database management system and application service system [6, 7]. The above research has laid the foundation for this paper.

Firstly, this paper analyzes and summarizes RBF, and then designs a logical structure framework that meets the actual needs and operation requirements and meets the user functions according to the specific situation to realize the systematic management of the platform, build the operation environment and improve the maintenance work. Finally, through the case study method, it verifies the feasibility of its application in English teaching, And put forward some feasible schemes to improve its development prospects.

2 Discussion on the Informatization of Constructive English Teaching Platform Based on RBF Algorithm

2.1 Constructive English Teaching

2.1.1 English Teaching Objectives

English learning is a complex process. Students have a certain cognitive foundation in English knowledge and skills. In teaching, we should have clear and specific training objectives. The main goal of English learning is to improve students' ability to use knowledge, obtain information and solve practical problems in language learning, help them establish a lifelong development consciousness, establish a correct and solid outlook on life and develop good living habits [8, 9].

2.1.2 Connotation of Constructivist Theoretical Learning View

There are many schools of constructivism, and there are many different views, but the basic views of cognition and learning are basically the same. In terms of learning, constructivists have the following views:

Active constructiveness of learning: learning is not a process in which learners simply mechanically and passively absorb external knowledge, but a process in which learners actively understand and create learners' internal and external experience. Learning is a process in which learners actively acquire knowledge. In the high-level problem-solving brainstorming activities, learners should constantly obtain the information related to the ongoing problem-solving activities, and constantly activate the original knowledge and experience to explain the current related phenomena. In such activities, we can give full

play to the interaction between old and new knowledge and experience, and it is an ideal medium for building knowledge.

Social interaction of learning: learning is a process in which learners internalize relevant knowledge and skills and gradually master relevant tools in the process of participating in a specific social culture. This process is usually completed in the collaborative interaction of the learning community. Constructivists believe that human innate potential is the basic condition of language learning, and social and cultural interaction is the key condition of human language learning.

Situational Learning: situations are always specific and changing. Learning the abstract concepts and rules of “de contextualization” cannot make a fair judgment on a specific and changing situation. Therefore, it is often difficult for learners to flexibly and effectively use the knowledge learned in school to solve practical problems in real life, and it is difficult for learners to effectively participate in life activities and social activities. Classroom design should pay attention to the important role of direct experience in students’ learning process, and pay attention to improving students’ independent exploration ability and self-monitoring ability.

2.2 Informatization of English Teaching

Information-based teaching means that teachers digitize teaching resources through computer network to enable students to obtain more timely and accurate information. It can not only improve classroom efficiency and quality, but also promote the learning and comprehensive application of English professional knowledge. The process of English learning is a series of dynamic changes from information processing, knowledge construction to problem solving. It has the following characteristics: (1) real time. Teachers and students can know the courses and teaching progress of the school anytime and anywhere, while others can also obtain the latest online education resources or materials through the Internet for their reference. At the same time, it can also monitor the teaching process and feed back to students, teachers and parents in time, so as to make the relationship between teachers and students more harmonious and friendly. (2) Interactivity. After class, using electronic information, teachers and students can discuss the learned knowledge on the teaching platform at intervals [10, 11].

2.3 RBF Algorithm

Artificial neural network is an information processing system developed by simulating the structure and mechanism of human brain. By imitating the structure and some reaction mechanisms of human brain, artificial neural network shows some basic characteristics of human brain as follows:

- (1) Artificial neural network is connected by many simple processing units to form a highly nonlinear system, which can deal with large-scale parallel problems. The biggest feature of the system is high parallelism and the function of each processing unit is very simple, but the connection of many simple processing units makes the whole system have rich functional characteristics. The information storage of artificial neural network is mainly distributed. The information does not exist in a

unit of the whole system, but is distributed in the whole artificial neural network system. Depending on the mechanism and human brain, the necrosis of some brain cells does not affect the overall function of the brain. Because there are many kinds of information in the artificial neural network system, and these information are distributed in each unit of the whole system, that is, each unit of the system intelligently stores some information, so that the whole system has the following functions: the “association” when receiving external information. It can correctly identify the incomplete information initially received through “Lenovo” by relying on the information stored in the database, and correct the incomplete information according to “experience”.

- (2) Another important feature of artificial neural network is its strong self-learning ability. If the external conditions change, the artificial neural network can give people the desired results of the modified external stimulation - results through a series of artificial retraining. RBF algorithm is a network-based, object-oriented and ubiquitous processing model. It decomposes a complex problem into multiple sub-categories and intelligently analyzes these subcategories. The biggest disadvantage of the global approximation neural network model is that each adjustable parameter in the whole network model will affect the output results. One input and one output in the network have a connection weight adaptation to them. Therefore, the learning speed of the whole network model is very slow and easy to fall into local minimization, which not only affects the accuracy, but also takes a long time. The difference of RBF neural network model is that it is a local approximation neural network model. It only generates connection weights for a local area of input data and affects the output. In other words, it affects each input and output. RBF neural network only needs to adjust several connection weights. Unlike BP, for each input and output data, The connection weight of the whole neural network is adjusted. This feature of RBF accelerates its learning speed and avoids the problem of local minimization. Because of this characteristic, RBF neural network model has been applied to many fields, such as image processing, fault diagnosis, function approximation, prediction data and so on. At present, there are mainly two types of neural networks with radial basis function, namely regularized RBF neural network and generalized RBF neural network.

The operation idea of neural network model is from n -dimensional input space to one-dimensional output space. The vectors of these two spaces are mapped, that is, the input vectors in n -dimensional space are mapped to one-dimensional space and become output vectors. We assume that there are n input vectors x_n , $n = 1, 2, 3, \dots, n$ in an n -dimensional space. The corresponding values of these input vectors in one-dimensional output space are Y_n , $n = 1, 2, 3, \dots, n$. In this case, such n inputs and outputs constitute the basic training sample set of artificial neural network. The main purpose of the so-called interpolation is to determine a mapping function $f(x)$ to satisfy the condition of Eq. 1:

$$F(x^n) = y^n, n = 1, 2, 3, \dots, n \quad (1)$$

In the above formula, function f describes an interpolation surface. This interpolation surface must pass all training samples. In the radial basis function, the method to solve this interpolation problem is to replace the interpolation surface in the above formula with n basis functions. The data in all training samples will have a basis function corresponding to it. The form of each basis function is shown in formula 2:

$$\phi\left(\|X - X^N\|\right), n = 1, 2, 3, \dots, n \quad (2)$$

Where, the basis function φ is a nonlinear function, and the training data point x_n is φ . The independent variable of the basis function is the distance between the point X in the input space and the center X^N . Because the distance is radial, the function φ is called radial basis function.

3 Experiment

3.1 Requirements for Constructive English Platform

In the traditional teaching mode, students accept knowledge passively, and the teacher's lecture is blackboard writing. Although this method can make it easier for teachers and students to understand and communicate, there are also many problems. In the traditional teaching mode, teachers are completed by students independently, and it is impossible to help them solve problems and solve problems like information technology. In information-based teaching, both teachers and students need to share resources through the network. As a language, learners should not only be the recipients of knowledge, but also learn how to understand themselves with the information they have learned. Therefore, the constructive education platform is to meet the high-level needs of learners for knowledge.

3.2 Information System Structure of Constructive English Teaching Platform

The structure of English teaching system is shown in Fig. 1. It is mainly composed of learners, teaching resource database and network support. For the construction of learning environment, teachers establish various forms of classrooms on the platform for students to practice independently, and adjust the relevant course contents according to the actual situation. Set up some auxiliary areas in the class to help them complete the after-school test task. Explain or display video materials for English knowledge points on video resources. Through these platforms, teachers can prepare teaching resources, network support, application software and other information for each class in advance.

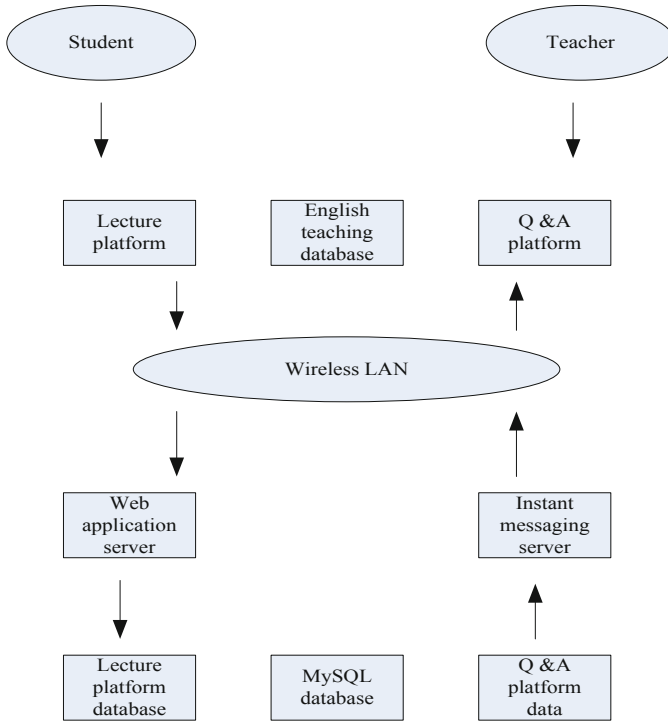


Fig. 1. System structure diagram

4 Discussion

4.1 Performance Test of English Teaching Platform

Table 1 is the platform operation performance test data.

Table 1. Test data

Number of tests	Processing connection maximum response time (ms)	Processing the connection minimum response time (ms)	Processing session maximum response time (ms)	Processing session minimum response time (ms)
1	0.741	0.523	2.357	1.385
2	1.521	0.957	1.473	1.140
3	1.745	1.024	1.854	1.247
4	1.574	1.042	1.037	0.782
5	2.476	1.254	2.387	1.574

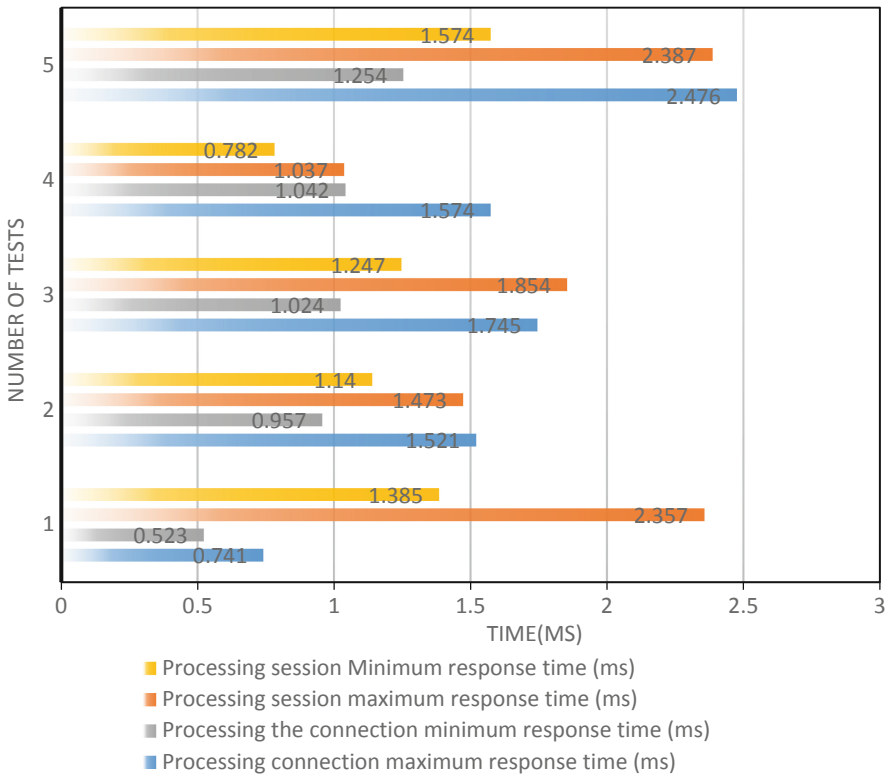


Fig. 2. System performance test

Figure 2 is the system operation performance test data chart. As can be seen from the above figure, the maximum response time and the minimum response time of the system when processing connections are about 1 ms. The maximum response time and minimum response time when processing sessions are maintained at about 1–2 ms. From the overall point of view of the system, this platform realizes the management of student information, course courseware and teacher users, and can effectively interact data in the network environment. Through the test, students can browse, modify and delete the platform information normally.

5 Conclusion

The main function of this platform system is to realize the sharing of teaching information resources through the communication between teachers and students, provide a good environment for English learning and improve the interaction between teachers and students. This paper mainly studies the learning platform information system based on RBF, by constructing a student-centered, teacher-led, teacher-student joint participation and interactive teaching mode.

Acknowledgements. This thesis is the result of the research project of foreign language education and teaching in vocational colleges in 2020. The name of the project is the in-depth integration of information technology and higher vocational English teaching in the new era (project number: WYJZW-2020-1377).

References

1. Li, G., Wang, F.: Research on art innovation teaching platform based on data mining algorithm. *Clust. Comput.* **22**(6), 13867–13872 (2018). <https://doi.org/10.1007/s10586-018-2119-x>
2. Shuli, Z.: Designing and practice of a college English teaching platform based on artificial intelligence. *J. Comput. Theor. Nanosci.* **14**(1), 104–108 (2017)
3. Mingxia, C., Han, Z., Yifeng, Z.: Research on temperature control algorithm of automatic packaging machine based on RBF neural network. *Packag. Eng.* **39**(19), 150–156 (2018)
4. Tsai, T.H., Hsu, C.W.: Implementation of fall detection system based on 3D skeleton for deep learning technique. *IEEE Access* **7**, 1 (2019)
5. Tomaszewski, D., Rapinski, J., Pelc-Mieczkowska, R.: Concept of AHRS algorithm designed for platform independent IMU attitude alignment. *Rep. Geodesy Geoinf.* **104**(1), 33–47 (2017)
6. Dixit, M.M., Vijaya, C.: DSP implementation of modified variable vector quantization based image compression using DCT and synthesis on FPGA. *Int. J. Inf. Technol.* **11**(2), 203–212 (2018)
7. Altmiller, G., Deal, B., Ebersole, N., et al.: Constructive feedback teaching strategy: a multisite study of its effectiveness. *Nurs. Educ. Perspect.* **39**(5), 291–296 (2018)
8. Agrawal, S.R., Mittal, D.: Constructive usage of WhatsApp in education sector for strengthening relations. *Int. J. Educ. Manag.* **33**(5), 00 (2019)
9. Ruegg, R., Williams, C.: [English Language Education] Teaching English for Academic Purposes (EAP) in Japan. Conclusion (Chapter10). **14**, 183–192 (2018). <https://doi.org/10.1007/978-981-10-8264-1>
10. Kokhan, R., Matsevko-Bekerska, L., Lysanets, Y.: Conversation analysis tool as an effective means for teaching the university courses of English and world literature. *Arab World English J.* **11**(4), 307–318 (2020)
11. Yanyao, H.: Construction of situational teaching system in commercial English courses. *Educ. Prog.* **007**(004), 137–141 (2017)



Dynamic Grouping Design System of Leisure Sports Based on Computer Intelligent Algorithm

Lipeng Zhang^(✉)

Nanchang Institute of Science and Technology, Nanchang, Jiangxi, China

Abstract. With the development of economy and the improvement of people's living standards, leisure sports have developed rapidly in China. As a sports project with strong interest, high participation, high flexibility and certain ornamental value, leisure sports have attracted more and more attention and gradually become an important part of the national fitness plan. Therefore, it is of great significance to study the dynamic grouping design of leisure sports based on computer intelligent algorithm. This paper first introduces the background and significance of leisure sports dynamic selection, and then expounds the requirements of dynamic grouping design. According to the application of computer intelligent algorithm, a dynamic grouping system based on computer intelligent algorithm is designed. Finally, the model is tested to study the possible errors in the process of dynamic grouping. The test results show that the random matching result error generated by the system in dynamic grouping is very small, which shows that it can meet the needs of users.

Keywords: Computer intelligent algorithm · Leisure sports · Dynamic grouping · System design

1 Introduction

Compared with the developed countries in the world, there is still a certain gap in the development level of China's sports field, and there are still many problems in competitive events. One outstanding performance is that athletes lack knowledge and theory of leisure technology [1, 2]. Under the traditional teaching mode, students are often unable to learn independently, and the intelligent and virtual networking system can effectively help them master sports skills and realize information exchange and sharing faster. By constructing a dynamic grouping model to improve the teaching efficiency of physical education teachers, it provides new ideas and methods for the development of China's physical education field [3, 4].

Many scholars have done relevant research on computer intelligence algorithms. The research and development of intelligent technology abroad has been quite mature and gradually applied to various fields. Among them, the developed countries represented by the United States and Germany have applied the sports virtual networking system earlier [5, 6]. Some domestic scientific research institutions and universities have also carried out in-depth research as an important topic. The relevant literature in China mainly

focuses on intelligent grouping algorithm. For example, some scholars have proposed a motion control strategy based on neural network technology. Other scholars analyzed and discussed how to realize interactive cooperation between intelligent devices from the perspective of dynamic and static [7, 8]. The above research has laid the foundation for this paper.

Starting with the development status of leisure sports, this paper analyzes the development and existing problems of leisure sports in China, and puts forward intelligent and personalized design. Mainly through the research of intelligent algorithm, combined with the relevant theoretical knowledge of computer technology, network communication and artificial intelligence, the system is designed.

2 Discussion on Dynamic Grouping Design of Leisure Sports Based on Computer Intelligent Algorithm

2.1 Leisure Sports

The term “leisure sports” derives from “leisure”. The word “leisure” appeared earlier. Although people’s understanding of leisure has a long history, so far, it has not formed an accurate and unified definition. Under the influence of different times and cultures, the understanding of this concept is also different. The word leisure in English is expressed as leisure, which comes from “licere” in Latin and the French noun “loisir” evolved from it [9, 10]. It is not difficult for us to see that the original meaning of leisure contains the meaning of freedom and no coercion. In Chinese, “Xiu” means “Jiqing” in Kangxi dictionary, and modern Chinese dictionary also has a similar meaning. “Leisure” generally refers to the scope. More often, it refers to the fields of morality and law. Therefore, from the word meaning, leisure expresses the dialectical relationship between work and rest, and is a spiritual activity. Leisure sports is a kind of physical activity, which can improve people’s form and reduce unnecessary pressure. Leisure sports refers to people-oriented activities that provide entertainment and services for the development of individual physical and mental health. It has two remarkable characteristics: one is to emphasize participation. Second, pay attention to personalization. It is not only a way of fitness and exercise, but more importantly, it can meet people’s physical and mental needs.

2.2 Dynamic Grouping Design

In the process of dynamic grouping, because users have high time requirements, there must be a real-time control mechanism to ensure that the data can be analyzed and processed accurately and timely when the system is running. The principle of dynamic grouping is to use computer technology to transmit random signals to intelligent nodes through the network, and then process and analyze the data. The basic idea is to divide randomness and aperiodicity into several different groups, adjust them according to their actual situation, analyze the problems or deficiencies in each group and give corresponding measures. In the process of dynamic grouping, we must consider the problems of randomness, unpredictability and complexity. In this process, a certain number and time

interval need to be set to ensure the accurate operation of the system. The classification needs to be divided according to the leisure sports time period used by different users. At the same time, combined with the improvement and perfection of the sports information collection method on the current smartphone software development platform, it is divided into dynamic grouping, that is, random grouping within a time period [11, 12].

In the traditional grouping control, a fixed time interval is usually used to judge and determine the randomness. But this method has some defects: first, when random factors change, it is easy to produce confusion. Second, because people are social animals, the interference can not be predicted. Third, non-standard actors compete and influence each other, resulting in unpredictable situations. Fourth, it is impossible to ensure the stable operation of the system, so the dynamic grouping must be controlled by means of non synchronization and high controllability. First of all, it is necessary to determine the mutual restriction relationship between various factors in the system according to the actual situation. Then the optimal scheme is obtained by analyzing each constraint condition. For two combinations of different constraint values, if an optimization function is 1, the optimal solution can be obtained, otherwise, the optimal result can not be obtained, and a new feasible function needs to be redesigned to meet the requirements of the algorithm and achieve the final goal.

2.3 Computer Intelligent Algorithm

2.3.1 Concept

Computer intelligent algorithm is a new technology based on traditional grouping technology and virtual networking. It takes computational reasoning as the core, and uses knowledge representation layer, rule analysis layer and hybrid discrete random process to build a dynamic grouping system. It has high predictability, real-time performance and robustness. It can simulate the random behavior generated when there are certain rules and uncertain factors in the thinking process of human brain to process complex system information. It can be applied to the field of sports. The intelligent algorithm realizes the optimal control by comparing and analyzing the user input information and output results. When there is a large difference in data or there is a certain deviation in the output signal, it will automatically judge whether the person is in the best state. If there is no error within the set time, the working mode of the system will be adjusted according to the current actual situation to promote the optimization of the whole motion process.

2.3.2 Algorithm Introduction

The process of particle swarm optimization algorithm is as follows: firstly, multi particle swarm optimization is carried out, and then the optimal solution is obtained iteratively. In the iterative process, follow two extreme values, the global extreme value $gbest$ and the individual extreme value $pbest$, and finally find the optimal solution. The speed update formula and position update formula 1 and 2 are shown.

$$V_{K+1} = v_k + c_1 r_1 (pbest_k - x_k) + c_2 r_2 (gbest_k - x_k) \quad (1)$$

$$x_{k+1} = x_k + v_{k+1} \quad (2)$$

In the formula, XK is the position of the current particle; VK is the velocity vector of the particle; $Gbest_k$ is the location of the optimal solution currently found by the whole population.

3 Experiment

3.1 System Objectives

The system applies the virtual sports dynamic grouping technology to the process of leisure sports, and realizes the analysis, prediction and judgment of random number sequence through intelligent algorithm. The research goal is leisure sports dynamic grouping technology. Using computer network can quickly, efficiently and accurately process data, improve work efficiency and prediction ability, and improve the problems of many complex redundant nodes, large path dependence, lack of self-organization ability and self-management in the previous manual division methods. At the same time, the application of the system can provide users with rich and diverse information acquisition platforms and good visibility operation. Through real-time monitoring of users and comparing the feedback information with the actual motion state, it can provide effective, accurate and timely data support for decision makers.

3.2 System Framework

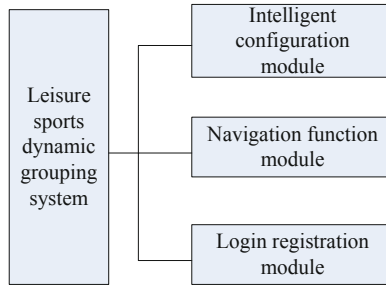


Fig. 1. System module composition

In this system, the main design idea is to group based on computer intelligent algorithm, and achieve the prediction goal through the data interaction between users and the information exchange between different types of user groups. From Fig. 1, the system is mainly composed of three modules: intelligent configuration, navigation function and login registration module. (1) The core of intelligent dispatching management includes user information collection and processing unit, data analysis and mining, etc. (2) Navigation, positioning and display interface: dynamic grouping settings are made according

to different sports types, which are set to query sports information and select positions in multiple channels. At the same time, the nearest and relatively close crowd in the corresponding area after reaching the designated place within the required time period is taken as the target object through the mobile terminal. (3) Login registration module: in this paper, we need to enter the account password information to enter the login page. When you click the “confirm” button, the system will automatically jump to the corresponding prompt state, and give specific answer options and corresponding time reminder values.

3.3 System Experiment Parameter Setting

Before the experiment, we first simulate the selected leisure sports dynamic grouping, first set the relationship between the parameter value and the location of the test point, and then start the next step. The randomly generated data is imported into the running environment through the database and state space analysis module in the system, and the degree of cooperation between the intelligent algorithm nodes in the corresponding time period between the corresponding data is calculated. Then, according to different strategies given by the system, the performance advantages and disadvantages of the experimental scheme are compared. In the process of setting experimental parameters, the random error and the influence of various factors on the results are mainly considered. In order to better realize the simulation effect and accuracy, the corresponding relationship table is established to represent the correlation between the two random variables. Finally, all possible problems in the functional module of the experimental system are programmed with MATLAB and set to the minimum error. On this basis, the instability of the maximum initial value of the algorithm is solved by designing a function A series of problems such as unreasonable peak flow distribution.

4 Discussion

4.1 System Performance Test and Analysis

Table 1 is the performance test data of leisure sports dynamic grouping design system based on computer intelligent algorithm.

Table 1. System performance test data

Number of tests	Test the estimated value	Actual test value	Error value	Error rate
1	253	258	5	0.019
2	236	234	2	0.008
3	257	254	3	0.011
4	241	240	1	0.004
5	296	287	9	0.030

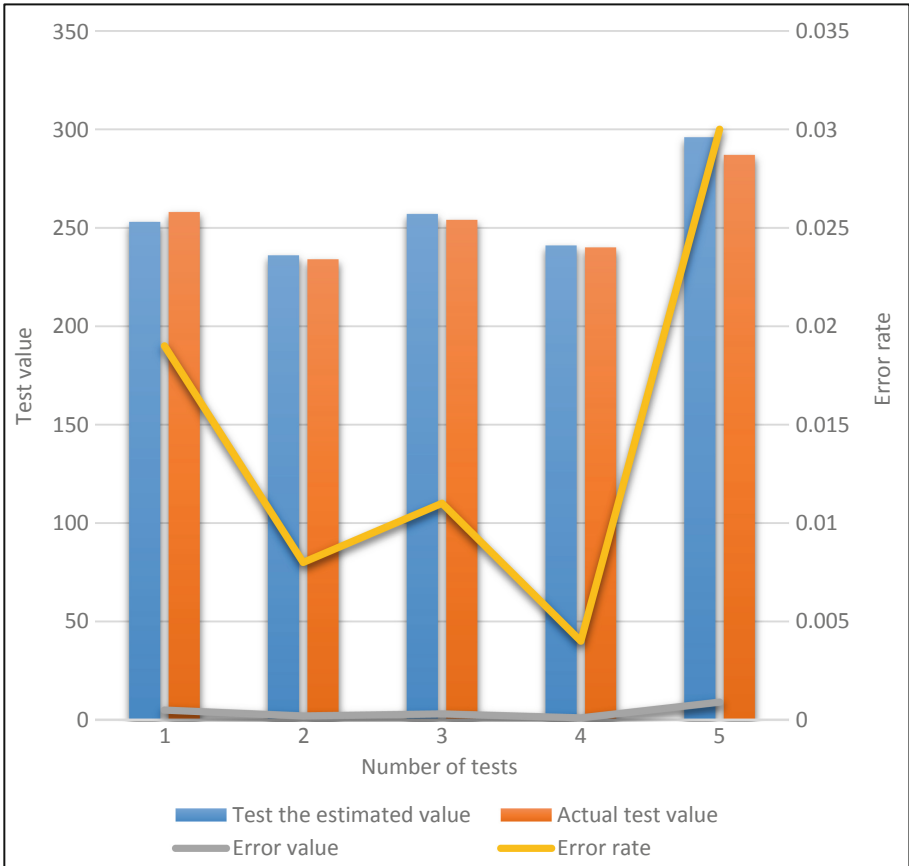


Fig. 2. System test error values for comparison

In the test process, the system error mainly comes from two aspects. One is the error caused by hardware factors. The second is the systematic error caused by improper human operation or other reasons. Software factors refer to the defects and deficiencies caused by failing to consider various problems in design and development, as well as the incompleteness caused by debugging and modification of the program, or the high degree of impact on the damaging results. As can be seen from Fig. 2, the result of random matching generated by the system during dynamic grouping is very small, which shows that it can meet the needs of users.

5 Conclusion

The design of this system is mainly to dynamically group leisure sports. In the traditional random process, there are large errors due to everyone's time, place and other factors. This paper puts forward a dynamic scheduling strategy to solve this problem. Based on

computer intelligent algorithm, this paper develops the mutual switching between real-time and non artificial types. Secondly, it realizes intelligent control by constructing network model. Finally, it establishes a complete system structure diagram by using virtual reality technology, so as to provide a certain degree of guidance for leisure sports in operation.

References

1. Pan, S.: Design of intelligent robot control system based on human–computer interaction. *Int. J. Syst. Assur. Eng. Manag.* 1–10 (2021)
2. Zhao, W., Luo, J., Li, S., Qi, J., Meng, H., Li, Y.: Design of dynamic calf weighing system based on moving-IIR filter algorithm. *J. Electr. Eng. Technol.* **16**(2), 1059–1069 (2020)
3. Zhang, Y., Wang, et al.: An anti-collision algorithm for RFID-based robots based on dynamic grouping binary trees. *Comput. Electr. Eng.* 2017, **63**, 91–98 (2017)
4. Kim, K., Kim, B., Lee, K., et al.: Design of integrated risk management-based dynamic driving control of automated vehicles. *IEEE Intell. Transp. Syst. Mag.* **9**(1), 57–73 (2017)
5. Huan, Z., Pengzhou, Z., Zeyang, G.: IEEE 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS) - Singapore (2018.6.6–2018.6.8). In: 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS) - K-means Text Dynamic Clustering Algorit., pp. 659–663 (2018)
6. Liu, H.: Establishment and maintenance of grass skiing resort turf based on climatic characteristics of Western Sichuan. *J. Landscape Res.* **9**(06), 7–9 (2017)
7. Nikitins, M.D., Griffin, P.A., Ch’Ng, S., et al.: A dynamic anatomical study of ulnar nerve motion after anterior transposition for cubital tunnel syndrome. *Hand Surg.* **7**(2), 177–82 (2019)
8. Zhang, L.: Design of a sports culture data fusion system based on a data mining algorithm. *Pers. Ubiquit. Comput.* **24**(1), 75–86 (2019)
9. Takacs, K., Drgona, P.: Design of precise simulation model of power converter of energy storage system located within micro-grids. *Transp. Res. Procedia* **55**, 904–910 (2021)
10. Liu, J., Chen, L., Rival, X.: Automatic verification of embedded system code manipulating dynamic structures stored in contiguous regions. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **37**(11), 2311–2322 (2018)
11. Yan, X., Gong, J., Wu, Q.: Pollution source intelligent location algorithm in water quality sensor networks. *Neural Comput. Appl.* **33**(1), 209–222 (2020)
12. Javier, D., Villar, J.R., De, L., et al.: Lecture notes in computer science. In: Hybrid Artificial Intelligent Systems. vol. 10870 || Orthogonal Learning Firefly Algorithm (Chapter 26), pp. 315–326 (2018). <https://doi.org/10.1007/978-3-319-92639-1>



The Frame Display of the Digital Clothing Display Design System Based on Intelligent Man-Machine Cooperation

Le Liu^(✉)

Shandong University of Art & Design, Jinan, Shandong, China

Abstract. In recent years, as people's requirements for economy and culture have become higher and higher, the demand for digital store displays has become increasingly urgent, and the research on clothing display design systems has also entered an unprecedented high period. The purpose of this article is to study the architecture display of the digital clothing display design system based on intelligent human-computer collaboration. The research status of clothing terminal display under clothing display and intelligent man-machine collaboration is described, and the functional requirements for the digital clothing display design system software are put forward. Taking the principle of system theory as another theoretical basis, the concept of clothing store system is proposed. Closely combining with the reality of clothing marketing, try to propose a quality evaluation system for clothing stores. An overall evaluation of the display of the clothing store system is made. The experimental results show that when the total score $F < 60$, the clothing store is a substandard structure display. In order to study and solve the complex display quality of the structure of the clothing display design system the problem provides an effective method.

Keywords: Human-computer collaboration · Digital design · Clothing display · Architectural display

1 Introduction

With the acceleration of my country's modernization and internationalization, people's requirements for high-tech information transmission technology, high-efficiency transmission methods, and high-quality transmission methods are becoming more and more urgent [1, 2]. Demonstration, as a behavior, is the expression of the original ecological beauty of biological vitality [3]. The display design is the sublimation and brilliant development of human conscious thought and behavior. After more than ten years of development, design works based on digital platform creation and dissemination have gradually formed their unique manifestations, internal characteristics and dissemination acceptance modes, and the expression methods of digital media design have developed towards multi-media integration and multi-dimensional development [4]. The mixing and superposition of image, text, video and sound will produce audiovisual effects different from graphic design, especially digital images, image processing technology and 3D animation technology provide more dynamic expression methods [5].

After hundreds of years of development, the ingenious design and unique charm of Nantong yarn-dyed earthen cloth in structure and pattern have been formed [6]. Roh J proposed a full-color complex holographic display system design, consisting of three R/G/B amplitude-only spatial light modulators and an achromatic Fourier filter. It is used for robust axial R/G/B color matching, so that the R/G/B holographic image light field can be three-dimensionally aligned. The synthesis algorithm for generating a full-color computer-generated hologram for the system is described, and the full-color optical reconstruction of the designed holographic three-dimensional image is proved through experiments [7]. Zhu X J proposed an interactive clothing design and a personalized virtual display with the user's own face. Taking suits as an example, this paper analyzes a customer interactive clothing design method based on genetic engineering ideas. Therefore, customers can rearrange clothing style elements, choose available colors and fabrics, and propose their own personalized suit styles. Based on Unity3D and VR technology, a Web 3D customized prototype system for personalized clothing has been developed. Combining the process of the system, the layout of structure and function is given. It also solves practical problems such as 3D face scanning, suit style design, fabric selection and accessories selection. Tests on the prototype system show that it can display realistic clothing and fabric effects, and provide users with an effective visual and customized experience [8].

This article has the following innovations: Combining the professional technology of clothing store design with computer intelligence technology, it is the first time that the intelligent design of clothing store based on knowledge engineering is proposed. Combining clothing display and display, store design knowledge with computer application technology, the software function is designed, and the information consulting function allows the system to flexibly and effectively read the information necessary for store design, making the system more targeted and intelligent. In the process of selecting the value of knowledge, not only the frequency of the use of the knowledge conclusion value, but also the number of continuous use are considered, so that the system can dynamically update the knowledge base according to the user's intention.

2 Research on the Frame Display of the Digital Clothing Display Design System Based on Intelligent Man-Machine Cooperation

2.1 Clothing Display

Display is to use some visual expressions and lighting techniques, including some supports and materials, to present information that will appear in public places, and have a greater impact on the audience's thoughts and behavior [9]. In order to achieve this goal, creative work was carried out. For example: overall design, layout design, display stand, color scheme, logo design, etc. In the design of clothing stores, display design is an important part of visual identity and a direct manifestation of clothing image. More and more brands are beginning to establish their brand image through terminal sales in various gate stores. Easy to publish and promote, it is not only widely accepted by operators, but also loved by customers [10, 11].

2.2 Apparel Terminal Display Under Smart Man-Machine Collaboration

In actual display design, the combination of several technologies is very important to improve the display effect and display landscape [12]. Effective control of multi-track, lighting, video playback time and method, and power changes are the key factors to create the best display effect. Among various control technologies, computer control is the most effective method, which can save a lot of manpower and time. On the basis of the computer control system, according to the requirements of the model, the intelligent computer system starts the corresponding music, lighting and video playback on time, and controls the intensity of the music and lighting according to the transmitted data. Through smart elements. Weaknesses, etc. save a lot of energy and protect the signal. In fact, these control technologies are not very cutting-edge technologies, but under the control of a computer system, the effects of multiple devices, such as fire and smoke, are combined to achieve event effects. Only by creatively applying these advanced technologies to clothing design can magical effects be achieved.

Now designers have begun to use scientific methods to analyze and determine the advantages and disadvantages of the physical and scientific environment of the exhibition space, and use computer technology to assist in design and drawing. The computer can accurately draw right angles and positional correlations, and present the visual images of the display space in a very detailed and realistic way. In many ways, the application of new technologies has greatly promoted the development of clothing store display design.

2.3 Software Function Requirements

2.3.1 Information Consultation

The system has a large-capacity, high-quality three-dimensional model database, which is updated regularly. Important information (mainly three-dimensional models) purchased from various relevant websites or institutions at home and abroad is stored in the corresponding database. Perform maintenance processing. Users can get the latest popular information at any time and download new shelves. At the same time, agents and retail management personnel in various places can inquire about the display specifications formulated by the company, the selling point information of the goods, the concept description of the store design and other information in the information database, which is convenient for display management and shopping guide training.

2.3.2 Space Design

Users can quickly design the store under the guidance of the store design wizard, or enter the store design module to design the store guide by themselves. According to the designer's positioning of the store, the system provides multiple series of reference stores, which users can choose to use according to their needs. Provide convenient tools to help users draw the basic space of the store, and plan the layout of the store (determine the flow direction of the store owner, the yin and sun faces, the clerk space, the goods space, and the configuration of the customer space). Users can choose from a large number of models stored in the database, drag and drop them in the custom store space, and carry

out display design. For the selected model, the user can change the size, material, add parts, recombination, etc.

2.3.3 Prop Design

The system provides lofting algorithm modeling function, users can make simple three-dimensional models such as shelves. At the same time, users can also modify and assemble existing shelves, including shelf design modification and editing, and individual modifications to parts.

2.3.4 Clothing Data Input and Goods Management

The software can call users' self-photographed clothing pictures and provide special module processing for display in stores. The user can input clothing information in the software and enter it into the database to facilitate the query and management of this batch of goods and the configuration of single-store goods.

2.3.5 Display

In the designed store, select clothing from the database for matching and display. In the design of display, the display design is often carried out from the perspective of only oneself, while ignoring the customer's feelings. This system provides a real-world roaming preview function to set up a walking route in the plan. After confirming, it will simulate the line of sight when walking along this route, walk in three-dimensional space freely and control the angle of view freely.

3 Research and Investigation on the Frame Display of the Digital Clothing Display Design System Based on Intelligent Man-Machine Cooperation

3.1 System Development Environment

Visual Basic was developed by Microsoft Corporation. Contains an event-based programming language that helps the development environment. It is widely used, easy to learn and use. The visual features of VB allow users to easily write a large amount of code to describe the interface, and simply drag and drop prefabricated objects into the interface to easily design the system interface. A database is a knowledge base for storing data. The data in the database is organized, described, and stored according to a specific data model, which has high data independence and scalability. This article chooses ADO data access as the way for VB to access database. ADO is a major development direction of Microsoft database technology. It has the advantages of simplicity and ease of use, no hierarchical structure, fast speed, low memory consumption, less disk space, and less code required.

3.2 Evaluation Index System

This article divides the rationality of clothing sales into 5 levels, and each level is assigned a score. The judges score item by item according to the degree to which the desired comment or requirement is met. The scores of individual indicators are shown in Formula 1:

$$F_{i-k} = a_j^* W_{i-k} \quad (1)$$

Among them, a_j is the grade score and W_{i-k} is the weight coefficient. The total score is shown in Formula 2:

$$F = \sum F_j \quad (2)$$

Among them, F_j is the score of the first-level indicator.

4 Research and Analysis of the Architecture Display of the Digital Clothing Display Design System Based on Intelligent Man-Machine Cooperation

4.1 Clothing Display Frame Display Design Module

Terminal display includes display products, colors, accessories, etc. The display of products adopts the design of general store space, cleverly combining unique design, ingenuity, beauty and charm. This series of products is completely unique in design and product design, and requires rich experience. This is the key for exhibition experts to solve complex and complex problems efficiently and with high quality.

If the system is to well present design ideas and practice expert thinking processes and techniques, the knowledge base must adopt and describe knowledge, rules, and experience in a balanced manner, as well as the experience of local problem-solving experts. The basic steps of shop display are: (1) Sub-region, product classification. (2) Perform display and display matching in accordance with display standards. (3) The appearance of the booth (pay attention to the theme, relevance and richness of the category. (4) Models wear themes to ensure that the styles of the models in the same group are consistent. (5) Adjust the focus accordingly, clarify the total area of the store and attract the attention of customers The top products are shown in Fig. 1:

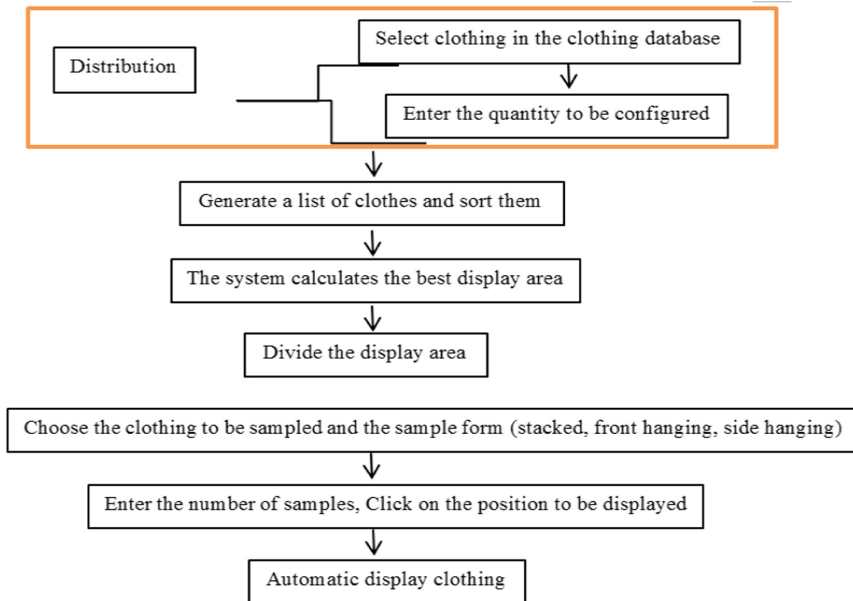


Fig. 1. Flow chart of display design

4.2 Evaluation of Digital Clothing Display Design System

Absolute evaluation, also called objective evaluation, is based on the purpose of evaluation, using a certain clothing store system as the basis for evaluation, that is, judging whether a $\geq A$ is established (where A is the expected goal). This evaluation is based on the quality of the evaluation object (the clothing store system).

Assuming the expected goal in advance, a score of 60 or less is considered a failure, a score of 60 to 80 is considered to be passing, and a score of 80 or more is considered to be excellent. Absolute evaluation: when the total score $F > 80$, the clothing store is an excellent structure display; when the total score $60 < F < 80$, the store is a qualified structure display; when the total score $F < 60$ pairs, the store is an unqualified structure display. A simulation evaluation of a certain store, as shown in Fig. 1. With a total score of 68.7, it is considered that the store is a qualified structure display as shown in Table 1 (Fig. 2).

For a specific clothing store, it is necessary to judge whether it is effective according to its store goal. If the goal of a clothing store is to make a profit, then as long as the output of the store (the sales profit of the store) is greater than the investment in development and maintenance, it is a profitable store. If the goal of a clothing store is an advertising image, then as long as it can help display products, promote brand and corporate image, it is an advertising store.

Table 1. General table of the quality evaluation index system of the clothing store system

Evaluation index	Single index	First level indicator
Window display	3.5	10.4
Shelf display	5.4	15.7
Booth display	3.4	7.5
People stand display	3.2	6.7
Cargo space configuration	4.3	8.6

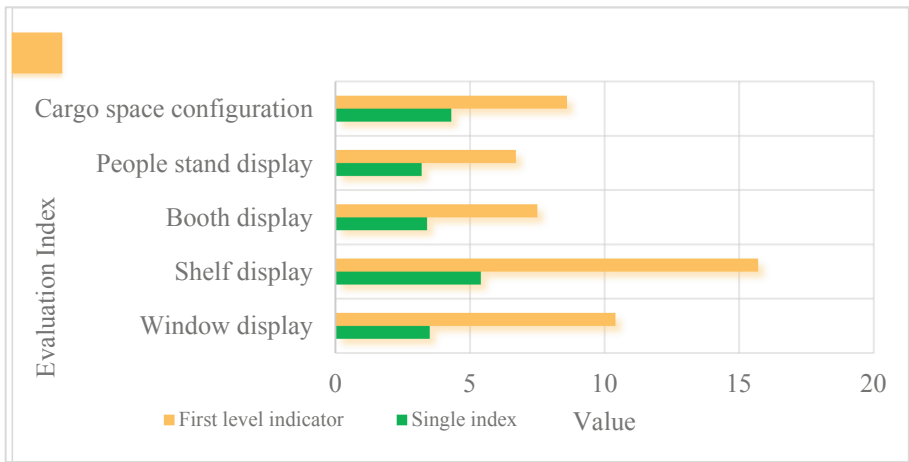


Fig. 2. System quality evaluation index of clothing store

5 Conclusions

With the improvement of the economic level, the people’s consumption consciousness is increasing, people are shopping more and more times, and people’s shopping environment has also undergone tremendous changes. And now shopping is not only about buying goods, but more about appreciating The form of beauty, the lifestyle of experiencing beauty, and the culture of experiencing beauty are a pleasant spiritual cultural journey. This paper uses the functional requirements of clothing terminal display software under intelligent human-machine cooperation to study clothing display from the actual point of view, which makes up for the lack of theoretical research on structural display. At the same time, this article summarizes the basic components of structural display by integrating literature and Created a digital clothing display design system evaluation, in order to provide the display personnel with the theoretical basis and practical application guidance of the clothing structure display.

References

1. Obeidat, A., Nabawi, H., Hashem, O., et al.: The impact of using interactive interior design on enhancing the performance of clothing shop. *J. Des. Sci. Appl. Arts* **1**(1), 146–153 (2020)
2. Stappers, P.J., Maldini, I.: The wardrobe as a system: exploring clothing consumption through design fiction. *J. Des. Res.* **17**(1), 3–25 (2019)
3. Choi, S., Kwon, S., Kim, H., et al.: Highly flexible and efficient fabric-based organic light-emitting devices for clothing-shaped wearable displays. *Sci. Rep.* **7**(1), 6424 (2017)
4. Wibowo, A.P., Sumiati, S., Donoriyanto, D.S.: Rancangan sistem informasi e-commerce untuk usaha fashion bountyclothing pada CV AI. *JUMINTEN* **1**(4), 24–36 (2020)
5. Ignatenko, G.: Clothing design and ornament function in the constructivist fashion of the 1920s–1930s. *Sci. Anal. J. Burganov House Space Cult.* **16**(2), 56–69 (2020)
6. Khisa, H., Hoque, M.E., Hanif, M.F., et al.: Impacts of visual merchandising on customer buying behavior in bangladeshi clothing stores. *Eur. J. Bus. Manag. Res.* **5**(6), 1–4 (2020)
7. Roh, J., Kim, K., Moon, E., et al.: Full-color holographic projection display system featuring an achromatic Fourier filter. *Opt. Exp.* **25**(13), 14774 (2017)
8. Babenko, L., Molev, M., Savelieva, N., et al.: Aspects of applying theory of human ecology to design safe heat-protective clothing for the disabled people. In: *E3S Web of Conferences*, vol. 175, no. 69, p. 15004 (2020)
9. Rosa, Y., Wahju, M.B.: Inventory and sales information system design on clothing store (case study: X fashion, Jakarta). *Jurnal Logistik Indonesia* **4**(1), 39–48 (2020)
10. Talbot, C.: Power, memory, architecture: contested sites on India's Deccan plateau, 1300–1600. *J. Roy. Asiatic Soc.* **27**(PT.1), 169–170 (2017)
11. Wirdyanto, S.H.: Cooling system for field service clothes Pt. Pln Persero based on Arduino Nano. *JEEE-U (J. Electr. Electron. Eng.-UMSIDA)* **4**(2), 121–130 (2020)
12. Landwehr, A., Browning, T., Jackson, C., et al.: An implementation of a packetized display protocol architecture for IRLED projector systems. *IEEE Photonics J.* **11**(2), 1–10 (2019)

Cloud Computing, Big Data, the Internet of Things, and Their System Intelligence



Classification and Recognition of Internet Literature Text Based on Data Mining Technology

Rong Xiong¹, Zichen Li², Yongjun Qi³(✉), and Thomas Lavoie⁴

¹ Faculty of Education, Guangdong Baiyun University, Guangzhou 510450, Guangdong, China

² Guangdong Polytechnic of Water Resources and Electric Engineering, Guangzhou 510635, Guangdong, China

³ Faculty of Megadate and Computing, Guangdong Baiyun University, Guangzhou 510450, Guangdong, China

⁴ University of Manitoba, Winnipeg, Canada

Abstract. With the rapid spread and development of the Internet, Internet text has also become the main source of information and receiving source in human daily life, and how to obtain more valuable information and knowledge from this information has also become an urgent problem to be solved. This article has carried out research on the classification and recognition of Internet literature text based on data mining technology. It understands the relevant theories of Internet literature text classification and recognition on the basis of literature data, and then designs the classification and recognition of Internet literature text based on data mining technology. In order to increase the accuracy of the analysis, the weighting method in classification and recognition was improved, and then tested to prove it. The result showed that the improved weighting method is more accurate in the same dimension than before the improvement, and it is in the four types of weighting. Among the methods, the accuracy is the largest and the classification effectiveness is better.

Keywords: Data mining · Network literature · Text classification · Weighting method

1 Introduction

As more and more users participate, the Internet has become like a giant spider's nest surrounding the world. Human communication is no longer limited by time and space. Traditional text and communication data storage methods are also migrating on the Internet [1, 2]. In this way, the text data on the Internet has become the largest data resource, and the text data on the Internet is growing in the form of geometric progression, which makes the Internet a treasure house of knowledge, containing huge useful information [3, 4]. On the one hand, more and more Internet users communicate and share through the Internet, creating more and more text data. These data come from all aspects of society and are closely related to human life and production. On the other hand, the growth rate

of these data far exceeds the capabilities of humans and computers [5, 6]. Text mining on the Internet is based on text mining, and text data on the Internet is used as the mining object, making it an important research application in the field of text mining [7].

According to the research results of document classification, some researchers have provided a semantic text graph model that combines vocabulary context and semantic information. This model uses actual words in the article as nodes in the graph, uses the positioning of words in sentences as directed edges, and uses word frequency signals such as nodes. Some researchers have also proposed methods for distributed expression of words. This type of word expression can effectively overcome the shortcomings and make similar or related words more similar to each other through word frequency calculation. Some researchers have proposed to use keywords as an important element of functional text to achieve text sorting tasks. The above-mentioned various methods are divided by traditional machine learning algorithms, and the performance of the sorting result depends to a large extent on the automatically generated text function, and the quality of the function directly determines the performance of the classification. In addition, it is not simple to adjust the modeling parameters, resulting in the entire process being time-consuming and laborious. Moreover, the way of artificially establishing features is based on a large amount of engineering domain knowledge, so the formed features are not fixed [8].

This paper studies the classification and recognition of network literature texts based on data mining technology. It analyzes the necessity of classification and recognition of network literature texts and the difficulties of network literature text classification and recognition on the basis of literature data, and then analyzes the network literature texts based on data mining technology. The classification recognition is designed, and the weighting method is improved in the design to improve the classification accuracy rate, and experiments are used to verify whether the improved method is effective [9, 10].

2 Research on Classification and Recognition of Online Literature Text

2.1 The Necessity of Text Classification and Recognition of Online Literature

The expression of user needs in the network environment is usually presented through the text network, which has many influences on the expression of user needs, such as the user's educational background, the user's knowledge level, the user's expressive ability, and expression needs. There are many factors. Text data mining technology is a good way to analyze user needs in the network environment, because there are many factors that affect the expression of user needs in the network environment, and it is impossible to manually identify these factors.

Text data mining technology is mainly to extract valuable information from complex text information. As an important field of semi-structured and unstructured data mining, text mining technology is compared with the existing popular user demand analysis technology. The application in the field of user demand analysis mainly includes four aspects of real-world reliable analysis results: fast acquisition of dynamic user demand, continuous acquisition of low-cost, real-time user demand information, and some unique advantages in collection.

- (1) The analysis result is true and reliable. Mainly because of reliable source of user demand and reliable text data mining technology. It is embodied in the equality and transparency of the Internet, so everyone is free to believe in real ideas. On the other hand, technological progress can eliminate useless information and unearth the most valuable and real user needs.
- (2) Quickly obtain the dynamic needs of users. With the rapid development of text data mining technology, not only can the web page receive the latest user demand information in real time, but it can also be processed and analyzed quickly to understand the dynamic needs of users.
- (3) Low-cost continuous acquisition. One is that the source of user needs is low in cost, and Internet information resources are shared, making it easy to collect user needs information for free. Second, the cost of technical analysis is low, and the existing technology is relatively mature.
- (4) Collect user demand information in real time. The dynamic nature of the Internet allows users to update their needs in real time. Companies can also use text-related technologies such as web crawlers to detect Internet-related information in real time and track user needs in a timely and accurate manner.

2.2 Difficulties in the Classification and Recognition of Online Literature Texts

(1) High-dimensional text objects

Even after pre-editing, the text set still contains thousands or hundreds of thousands of feature words, and the feature word dimension directly corresponds to the text table dimension. In view of the high dimensionality of text tables, common data mining methods are computationally expensive or costly, or some mining algorithms work well for low-dimensional data in terms of performance and output. Therefore, the existing methods for adapting to the word processing problem of high computational load and high resource consumption need to be improved, and new text representation methods or effective dimensional methods can be studied at the same time.

(2) The authenticity of the text object

When using the vector space model to represent text, each text is described by an attribute word contained in the entire text set. Since a text set usually contains many topics, each topic is represented by a different subset of feature words. Therefore, in this case, the non-zero elements of the text vector are usually very few, and the text representation is significantly diluted, because the attribute words contained in the text are only a small part of the attribute words in the entire text, so the difference between the text objects. The similarity is very low, close to 0, and the similarity tends to be the same. These have a significant negative impact on similarity-based clustering algorithms and have a direct impact on classification accuracy.

(3) Semantic issues

Chinese texts also have their own natural language phenomena, such as synonyms and synonyms. The phenomenon of synonyms and synonyms means that the same subject or content can be described in different ways. Their existence significantly reduces the accuracy and efficiency of text grouping. By using the latent semantic index method to edit text objects, the purpose of reducing the dimension of feature words and strengthening semantic relevance can be achieved.

(4) Complicated description problem

The cluster description aims to explain the results of the grouping, making it easier for users to understand the results of the clustering, and allowing them to quickly see whether the document categories they created are relevant.

2.3 Classification Algorithm Improvement

The TF-IDF weight combines the absolute word frequency method and the inverse document frequency method to correct the incorrect distribution between the category of the ignored word frequency method attribute and the frequency of the IDF ignored attribute in the text. In fact, it is simple and easy to understand. It can be seen from formula (1) that the value of TF-IDF is directly proportional to the number of times the attribute element appears in the document, and inversely proportional to the number of times the attribute element appears in the document set. However, its disadvantage is that it ignores the distribution of data and the position of attribute elements in the class document.

$$w_{ij} = tf_{ij} \times \log\left(\frac{N}{n_i}\right) \tag{1}$$

Variable description: N is the total number of texts, 1 is the weight of the feature item t appearing in document d, 2 is the word frequency, and 3 is the number of documents in the training set that contain t.

Due to the shortcomings of TF-IDF, the document frequency needs to be added to the TF-IDF Eq. (1) to change the weighting method of the original TF-IDF. It creates a frequency factor for the document. The formula for the reflection coefficient α and the formula for the end class coefficient b are as follows.

$$\alpha(t_j|C_i) = \frac{f_{ij}}{f_i} \tag{2}$$

$$\beta(t_j|C_i) = \frac{f_u}{n_i} \tag{3}$$

Where 1 is the number of documents with feature t in category 2, and 3 is the number of texts with t in the training set.

3 Network Literature Text Classification and Recognition Based on Data Mining Technology

3.1 Schematic Diagram of Text Classification Process

The text classification in this article has two main stages. One is the training process, and the other is the testing process (the process of classifying unknown text). The training stage is the process of learning classifiers, and the formed classifiers are used to test unknown categories of text in the next stage, and then the classification results are analyzed through the evaluation model. The preprocessing steps include word segmentation, end word deletion, attribute weighting, attribute selection and vector representation. The overall text sorting box is shown in Fig. 1.

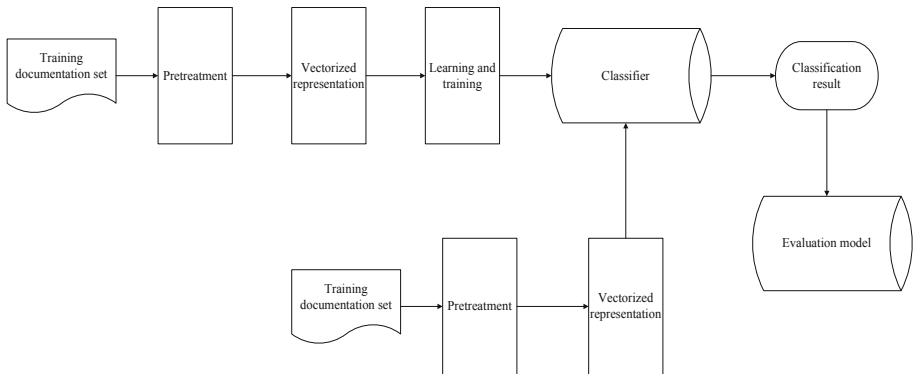


Fig. 1. Schematic diagram of the overall process of text classification

3.2 Classification Model

Random forest is an ensemble model that uses a general classification tree (decision tree) as the basic classifier. In this classification method, the data sample is randomly modified many times, and some variables are randomly selected from the sample data. Since the sorting tree is created for each node of the sorting tree, the samples are randomly generated, and the node variables are also randomly generated, so the class tree of each category can grow without pruning. Finally, vote for each data point through these sorting trees. Random forest is an ensemble model based on classified data. The advantage of the ensemble model is that it can effectively solve the problem of unbalanced data classification. Therefore, the classification model of this article chooses a random forest.

3.3 Data Dimensionality Reduction

There are many ways to reduce dimensionality, but the most common one is feature extraction. Feature extraction is actually to map and transform the original data set, and map the original data in the high-dimensional space to the low-dimensional space to reduce the loss of information after the transformation. This document uses the singular value decomposition (SVD) method to reduce the size of the original data.

4 Detection of Classification Algorithms

4.1 Data Selection

The data selected in this article was developed by the natural language processing team of Fudan University, which is widely used in Chinese text classification research. In this work, 15 categories of news data sets with few documents will be deleted, and 8 categories will be selected for investigation. That is, art, space, computer, environment, agriculture, economy, politics, sports, a total of 9090 texts.

4.2 Experimental Design

Validation experiments are performed on the same classifier KNN to verify the effectiveness of the improved feature weighting method TF-IDF-C based on between-class and intra-class factors in this chapter. Table 1 shows the experimental data of using the KNN classifier to use four weighting methods (Boolean, TF, TF-IDF and TF-IDF-C) on various feature dimensions.

Table 1. KNN classifier uses four weighting methods (Boolean, TF, TF-IDF and TF-IDF-C) on various feature dimensions of classification experiment results

	300	800	1300	1800	2300	2800
Boolean	0.74	0.69	0.67	0.62	0.57	0.55
TF	0.81	0.82	0.82	0.82	0.81	0.79
TF-IDF	0.82	0.81	0.79	0.78	0.79	0.78
TF-IDF-C	0.85	0.88	0.89	0.89	0.88	0.88

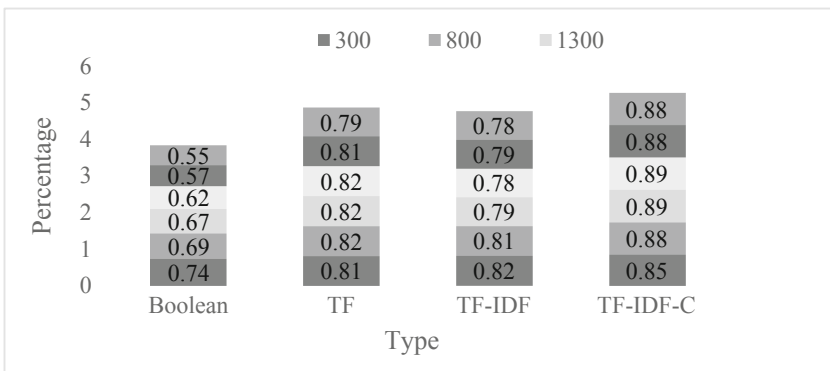


Fig. 2. KNN classifier uses four weighting methods (Boolean, TF, TF-IDF and TF-IDF-C) on various feature dimensions of classification experiment results

It can be seen from Fig. 2 that for the Boolean weighting method, the larger the feature size, the lower the accuracy, and the larger the feature size, the accuracy of other weighting methods is slightly higher, but due to the accuracy of TF, TF-IDF, and TF-IDF-C There is no significant improvement since 1300 dimensions, so it is recommended to choose a dimension around 1500. This preliminarily confirms the superiority of the improved TF-IDF-C weighting method, which is superior to the accuracy level before the improvement, is the most accurate of the four weighting methods, and has excellent classification results.

The following three indicators are used to compare the four weighting methods in Table 1 to obtain Table 2. It is further found that the improved weighting method TF-IDF-C has a good classification effect in most of the eight categories.

Table 2. Evaluation index value of each category of four weighting methods

Weighting method	Index	1	2	3	4	5	6	7	8
Boolean	P	0.76	0.41	0.91	0.87	0.84	0.91	0.51	0.85
	R	0.30	0.95	0.62	0.62	0.50	0.61	0.52	0.72
	Fl	0.44	0.54	0.72	0.73	0.63	0.73	0.60	0.66
TF-IDF	P	0.76	0.62	0.92	0.86	0.82	0.92	0.81	0.92
	R	0.57	0.93	0.83	0.76	0.81	0.81	0.73	0.80
	Fl	0.64	0.76	0.90	0.81	0.81	0.86	0.77	0.86
TF-IDF-C	P	0.94	0.79	0.98	0.88	0.97	0.88	0.98	0.97
	R	0.30	0.97	0.98	0.91	0.85	0.97	0.94	0.93
	Fl	0.45	0.88	0.98	0.89	0.87	0.91	0.94	0.95
TF	P	0.85	0.43	0.93	0.91	0.80	0.94	0.92	0.95
	R	0.43	1.00	0.90	0.64	0.74	0.76	0.74	0.72
	Fl	0.58	0.61	0.91	0.76	0.82	0.83	0.81	0.81

It can be seen from Fig. 3 that among the four weighting methods, the Boolean weighting method is the lowest in the second and seventh categories, and the TF weighting method is better than the Boolean weighting method. Compared with TF-IDF, TF-IDF-C has improved accuracy in most categories. The first and second categories have the largest increase of about 18%, followed by the fifth category, with an increase of about 10%. In the sixth category, it has dropped by about 10%. In general, the correct rate is higher than before the improvement.

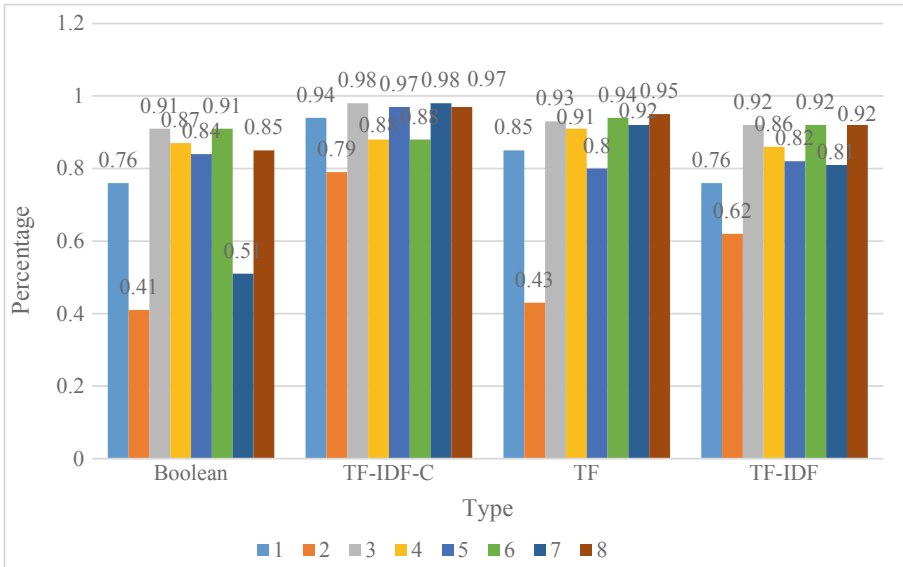


Fig. 3. Evaluation index value of each category of four weighting methods

5 Conclusions

This paper studies the classification and recognition of Internet literature texts based on data mining technology. After analyzing related theories, design the classification and recognition of Internet literature texts based on data mining technology, and improve the weighting method in the classification to improve. The performance of the text genre is then verified through experiments to improve the effectiveness of the method. The experimental results show that the accuracy of text classification is improved after the weighting method is improved in this paper.

Acknowledgements. Natural Science Foundation of Xinjiang Uygur Autonomous Region: 2020D01A34.

Education Reform Fund project of Guangdong Province: CXQX-JY201806.

Feature innovation project of colleges and universities in Guangdong province: 2020KTSCX163.

Feature innovation project of colleges and universities in Guangdong province: 2018KTSCX256.

Guangdong baiyun university key project: 2019BYKYZ02.

Special project in key fields of colleges and universities in Guangdong Province: 2020ZDZX3009.

References

1. Li, Z.: Implementation of classification and recognition algorithm for text information based on support vector machine. *Int. J. Pattern Recognit. Artif. Intell.* **34**(08), 360–372 (2020)

2. Bakkali, S., Ming, Z., Coustaty, M., Rusiol, M.: Visual and textual deep feature fusion for document image classification. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE (2020)
3. Zhao, M., Wang, R.Q., Yin, F., Zhang, X.Y., Ogier, J.M.: Fast text/non-text image classification with knowledge distillation. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE (2019)
4. Bo, T., Kay, S., He, H.: Toward optimal feature selection in naive bayes for text categorization. *IEEE Trans. Knowl. Data Eng.* **28**(9), 2508–2521 (2016)
5. Elghazel, H., Aussem, A., Gharroudi, O., et al.: Ensemble multi-label text categorization based on rotation forest and latent semantic indexing. *Expert Syst. Appl.* **57**(Sep), 1–11 (2016)
6. Jiang, M., Pan, Z., Li, N.: Multi-label text categorization using L21-norm minimization extreme learning machine. *Neurocomputing* **261**(Oct.25), 4–10 (2017)
7. Naderalvojud, B., Sezer, E.A.: Term evaluation metrics in imbalanced text categorization. *Nat. Lang. Eng.* **26**(1), 1–17 (2019)
8. Rossi, R.G., Lopes, A.D.A., Rezende, S.O.: Using bipartite heterogeneous networks to speed up inductive semi-supervised learning and improve automatic text categorization. *Knowl.-Based Syst.* **132**(Sep.15), 94–118 (2017)
9. Wang, T., Cai, Y., Leung, H.-F., Lau, R.Y.K., Xie, H., Li, Q.: On entropy-based term weighting schemes for text categorization. *Knowl. Inf. Syst.* **63**(9), 2313–2346 (2021). <https://doi.org/10.1007/s10115-021-01581-5>
10. Belazzoug, M., Touahria, M., Nouioua, F., Brahimi, M.: An improved sine cosine algorithm to select features for text categorization. *J. King Saud Univ. Comput. Inf. Sci.* **32**(4), 454–464 (2020)



Design and Realization of Applied Linguistics MOOC System Based on Feature Extraction Algorithm

Hongyan Sang^(✉)

Science and Technology Department, Chongqing Vocational College of Transportation,
Chongqing 402260, China

Abstract. With the rapid development and popularization of modern information technology, many colleges and universities have used their own linguistic MOOC system, which has achieved remarkable results in management teaching, which has facilitated the learning of students. However, at that time, most systems were limited by the performance of software and hardware at the development and technical level, and did not consider how to develop and use the functions of the set-up system so that each student can learn easily. This article introduces the relevant theories of feature extraction algorithms, combined with the practical application of the university selection system, and applies the feature extraction algorithms to the MOOC system of applied linguistics. Analyze students' historical academic performance on various topics, help students choose courses with the most potential, avoid students blindly choosing courses, so that all students can have their own satisfactory learning resources. Experimental research shows that this paper chooses 0.6 as the early probability of feature extraction algorithm, which is conducive to the realization of this process from global search to local search.

Keywords: Feature extraction algorithm · Applied linguistics · MOOC system · New teaching mode

1 Introduction

Compared with the traditional teaching mode, the MOOC system based on feature extraction algorithm has its unique advantages. The first is the refinement of teaching resources: MOOC videos are generally meticulously arranged and repeatedly scrutinized, and can only be launched after multiple audits. The method and order of knowledge teaching are all selected to the best state, making users' learning efficiency the most [1, 2]. The second is the differentiation of the teaching process: In traditional classrooms, teachers can only use a single teaching method to deal with many different individual students, which will inevitably cause differences in teaching effects. In the learning process of MOOC, students can Pause and retreat at any time, think over and re-think unfamiliar knowledge points, and strengthen memory.

In the research on the design of applied linguistics MOOC system based on feature extraction algorithm, many scholars have studied it and achieved good results. For example, Shen Y proposed a course scheduling process based on Lagrangian algorithm, using The branch and bound method is used for solving and so on [3]. Aarti explored the issue of the scheduling of graduate students' timetables, and proposed that the scheduling of timetables should be arranged as a unit of human-level curriculum, and the way of resolving conflicts is consistent with the SAPHIR system [4]. It can be seen that the advantage of applying algorithms to teaching system education compared to traditional online education lies in the refinement of the content of the course, taking knowledge points as the teaching object, and adding the interaction and testing links between teaching and learning during the learning process. This is an ordinary online What education does not have, thus expanding the teaching methods, improving the quality of teaching, promoting students' learning attitude, and enhancing students' learning enthusiasm.

By comparing and researching the feature extraction algorithm with other algorithms, the mathematical model corresponding to the lesson scheduling problem is designed and constructed, the traditional feature extraction algorithm is improved and experimental verification is carried out, and the improved algorithm is applied to the final course selection in the system. In this paper, the experimental use cases are determined, and then the algorithm's convergence speed under different parameters is tested, and then the feature extraction algorithm is compared with the experiment, and the performance evaluation and result analysis are carried out.

2 Research on the Design of Applied Linguistics MOOC System Based on Feature Extraction Algorithm

2.1 Demand Analysis and Research on MOOC Teaching Platform of Applied Linguistics

(1) Analysis of user management needs

After investigation, research and analysis of the teachers and students of the college, the users in the MOOC teaching platform are finally divided into three roles, and have different role permissions, namely, administrator permissions, teaching teacher permissions, and online student permissions. This distribution method can not only ensure the normal operation of the MOOC platform, but also reasonably plan the college personnel who use the MOOC platform, ensuring the combination of personnel and the system [5, 6]. The administrator's authority is to manage the accounts of instructors and online students, including user management settings, login user authority settings, and system settings. In the platform system, it is the role of a super administrator. The authority of the instructor is the analysis and use of the teaching function and test results in the teaching and learning interactive module for the release and management of the lectured course video. The permission of online students is to choose their favorite courses and learn the selected courses and use the learning function in the teaching and learning interactive module.

(2) Demand analysis of course information management

Curriculum information management is mainly to conduct reasonable and standardized management of the released curriculum resources by the teachers of the

college, including the management of overall curriculum information and chapter information. Among them, the administrator has the authority to set the course professional catalog and course information, as well as the management authority of the course [7, 8]. The instructor has the authority to operate the course chapters and the authority to publish course videos and exercises. This authority is only for the courses published by the person, and cannot operate other courses that are not published by the person. Students have the authority to query course information and course learning, and do not have the authority to modify any operations. The administrator enters the course professional catalog. After the teacher logs in, he can select the professional category that has been entered, and add the released course video to the category. After logging in, the students can search through the category when they choose the course. Find through the information of the instructor. After the course is released, the instructor also has the function of modifying the released course information, including modification, deletion, and re-addition.

1) Course information release

The release of course information mainly refers to the new creation of MOOC courses by the instructor, including the introduction of the courses and the operation of the course news release function. Course message is the management of important information for related courses. The corresponding recipient of the message is all users, including teachers and students. The form of the message includes text format and hyperlink format [9, 10]. The instructor has the right to publish, modify and delete the relevant messages of the published course, and the students only have the right to view the news of the selected learning course.

2) Course management

Course management includes the operations of adding, modifying and deleting course chapters. The instructor only has the authority to perform the above operations on the courses published by him, and he cannot exceed his authority. The management of courses in the platform is operated by chapters as the basic unit. Course management implements a three-level management model, namely courses, course chapters, and course sections. Online videos and practice questions are categorized by chapter.

(3) Demand analysis of course video management

Course video management includes two functional modules: course video management and course interaction management. Course video management is based on course information management to manage course videos. The course videos are added, modified, and deleted according to chapters, and the courses are released for learners to study. The course video management module performs authority management. Only the instructor has the above authority for the course videos posted by him, while other users have only the authority to watch the video, which can protect the security of the video and prevent destructive operations [11, 12].

2.2 The Basic Idea and Algorithm of Feature Extraction Method Design

Calculated as follows:

$$a_{ij} = tf_{ij} \times \left(1 - \sum_j \frac{(tf_{ij}/gf_i) \log(tf_{ij}/gf_i)}{\log(N)}\right) \quad (1)$$

Among them, tf_{ij} and gf_i respectively represent the frequency of search term i in MOOC system j and the entire system set;

In order to eliminate the inconsistency in the number of words caused by the inconsistency of the text length and bring the inconsistency of discrimination to the weight evaluation, the author sets the weight a_{ij} . After normalization, the calculation formula is:

$$a_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^M a_{ij}^2}} \quad (2)$$

Among them, a_{ij} represents the weight of the i -th word in the j -th MOOC system, and M is the number of search words.

2.3 Requirement Analysis of the Course Examination System Based on Feature Extraction Algorithm

(1) Demand analysis of the system's volume subsystem

Traditional methods and test paper management methods have been unable to meet the development of this field in the new era. The demand analysis of the test paper composition subsystem of the feature extraction algorithm course examination system is as follows:

- 1) Teachers can manage the questions in the question bank accordingly, including display, delete, add, and modify;
- 2) Be able to set various parameters of the new test paper to be composed, such as the overall difficulty, the scope and proportion of knowledge points or chapters;
- 3) Capable of keeping pictures and texts for the test questions to be entered into the system. Including: diagrams, formulas, etc.
- 4) The test questions can be edited separately to add, or unified import can be realized. When adding the question bank, each index of the test questions must be added;

(2) Architecture design of online applied linguistics MOOC examination system

Ajax layer: This layer performs processing similar to middleware, and then the "middleware" accesses the database to perform data storage operations. The data is returned immediately after the storage, so that asynchronous data can be transmitted between the browser and the Web server. This layer The results of database operations are returned to the top layer for asynchronous updates and displayed to users. It is a common technique for creating fast and dynamic web

pages, which can update part of the web page without reloading the entire web page. User layer: In this layer, it is responsible for displaying the interface to the user. The displayed data only needs to be beautiful and easy to use. The accuracy of the data is guaranteed by the user layer.

Research on the overall front-end and back-end technologies of the system. The jsp and php languages can be easily implemented and developed. The platform cooperates with the https protocol and user md5 data flow authentication to ensure high data encryption, so we choose to use the JAVA development language. The system adopts the B/S structure, the user interface is friendly, and the professional requirements for the final user are low. The operator can easily perform common operations on the question bank and test papers through the interface.

2.4 Build a Situational Simulation Support System

(1) Domain knowledge base

The domain knowledge base is based on the curriculum goals and the needs of the system to realize the situational simulation. The main function is to facilitate the retrieval and sharing of knowledge by teachers and students, and to improve the efficiency of knowledge retrieval. It is the basic component of the situational simulation support system. The domain knowledge base in the C3S system is composed of a knowledge system library and a teaching material library. The knowledge system library includes subject definitions, theories, formulas, etc., such as the theoretical knowledge of the journey problem in junior high school mathematics. The teaching material library is the relevant courseware corresponding to each knowledge point.

(2) Context creation module

The situation creation module is a component that reconstructs the problem situation and problem-solving situation when teachers and students are “teaching” and “learning”. Its main function is to create a simulation of courseware. Drawing lines, etc.) and simulation components, and can combine various components with each other. The user can set the attributes, events and actions of the component, and form and output a situational simulation file (XML file structure) after creating the component. The module is simple to operate and can better facilitate teachers to create courseware.

(3) Simulation analysis module

After the simulation analysis module creates scenarios for teachers and students, the system performs simulation analysis and execution on the scenario simulation files. The main function is to perform simulation analysis and execution programs on the scenario simulation files after the scenario creation, and output computer programs.

The MOOC system is based on the B/S structure. On the server side, it establishes a Webservice under the NET framework, and requests a simulation file (XML file) for receiving user learning resources.

(4) Situational awareness module

Context aware module (Context aware), this module is to better let teachers understand the behavior information of students in the process of restructuring problem situations and solving problem situations. The main function is to obtain

students' learning situation information, including learning Preferences, mouse clicks and movement trajectories, keyboard input, order of answering and other contextual information, the obtained students' preference behavior information is based on a period of time, not preference information at a certain point in time.

3 Research on Experimental Design of MOOC System Based on Feature Extraction Algorithm

3.1 The Purpose of the Experiment

Experiments on the MOOC course selection system based on feature extraction algorithms to verify the feasibility of the feature extraction algorithm to solve the scheduling problem, determine the best value of the algorithm parameters in the system, and the improved feature extraction algorithm for the efficiency of the course selection system the promotion.

3.2 Experimental Test Data

The classes to be scheduled are the first and second grades, a total of 21 classes, of which there are 8 subjects in each class in the first year, 10 subjects in each class in the sophomore year, and a total of 60 classes in the two years, arrange 23 teachers and 14 classrooms. Select some courses, teachers, and classrooms to set non-arrangeable courses and priority courses.

4 Experimental Analysis and Research of MOOC System Based on Feature Extraction Algorithm

4.1 Crossover Probability Experiment Analysis

The choice of crossover probability is mostly between 0.35–0.91. This article selects 0.4, 0.6, and 0.8 for experiments. The test case adopts experimental data one, and each group of experiments is run multiple times and the running time of 5 times is randomly recorded. The results obtained are shown in Table 1:

Table 1. Comparison of running time with different crossover probabilities

Experiment number	Crossover probability 0.8 (s)	Crossover probability 0.6 (s)	Cross probability 0.4 (s)
1	25.73	31.50	34.16
2	27.91	33.18	36.99
3	25.16	29.71	36.12
4	27.33	31.26	37.83
5	24.58	32.72	36.75

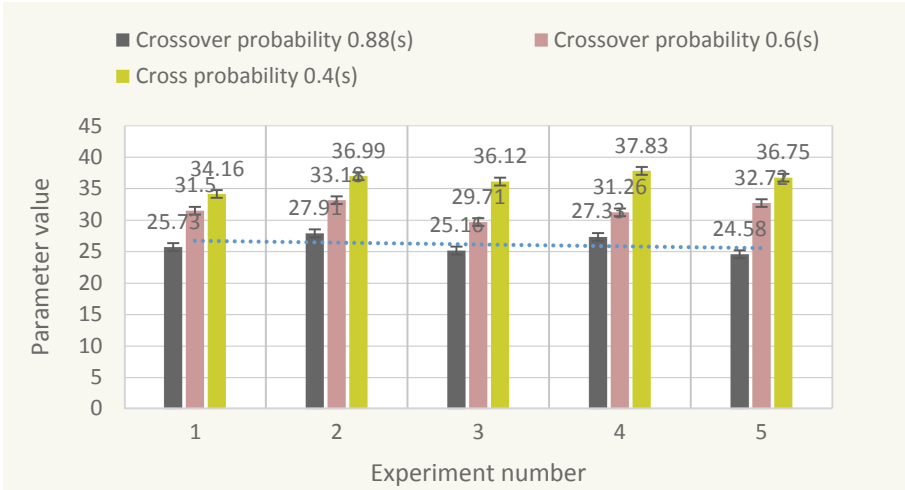


Fig. 1. Comparison of running time with different crossover probabilities

It can be seen from Fig. 1 that the greater the crossover probability, the shorter the program running time, and the higher the execution efficiency of the algorithm, but in order to prevent premature convergence, it cannot be selected too much. The effect of crossover probability on efficiency is not very obvious. After many tests and comparisons, this paper selects 0.6 as the early probability of the algorithm, and 0.6 as the mid-late probability of the algorithm.

4.2 Experimental Analysis of Mutation Probability

This paper selects 0.01, 0.05, and 0.09 for corresponding experiments. The test case uses experimental data one, and each group of experiments is run multiple times and the running time is randomly recorded 3 times. The results obtained are shown in Table 2:

Table 2. Comparison of running time with different mutation probabilities

Experiment number	Probability of mutation 0.02 (s)	Probability of mutation 0.04 (s)	Probability of mutation 0.06 (s)
1	24.61	30.15	37.99
2	25.67	31.29	38.52
3	27.31	32.47	37.69
4	26.82	30.18	36.84
5	26.73	31.19	35.75

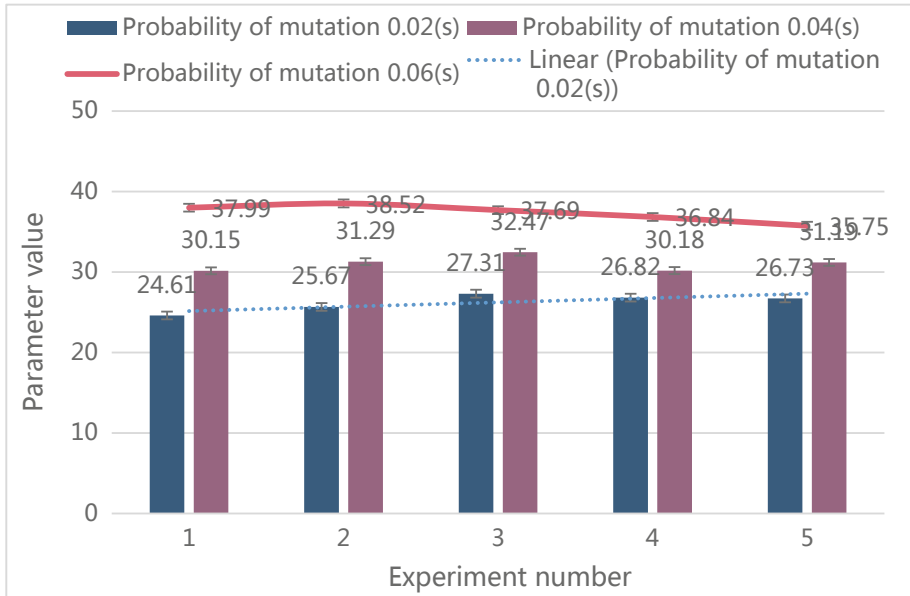


Fig. 2. Comparison of running time with different mutation probabilities

As shown in Fig. 2, it can be concluded that in the early stage of the algorithm operation, the fitness of the individual is generally low, so the degree of search must be increased. At this time, a larger crossover probability should be selected and a smaller mutation probability should be selected to make the algorithm more effective. The population fitness has been rapidly improved.

5 Conclusions

The research on the applied linguistics MOOC examination system based on the feature extraction algorithm is of positive significance. This research is a new breakthrough in the algorithm of the traditional question-making software system. The specific advantage of the new algorithm is that the stability is better. The convergence speed is fast, the performance is also good, and the teacher's labor is saved. The algorithm in this paper can not only play a very good research significance and guiding significance for training assessment, but also can be applied to other professional course assessment and assessment, so as to improve the efficiency and quality of teaching assessment.

References

1. Shaaban, S.: Groundwater assessment using feature extraction algorithm combined with complex proportional assessment method and standard deviation. *Int. J. Intell. Eng. Syst.* **14**(2), 306–313 (2021)
2. Li, B., Mo, D., Song, Z., et al.: Feature extraction algorithm of precession target based on image length and Doppler broadening. *Appl. Opt.* **59**(29), 9098–9103 (2020)

3. Shen, Y., Zheng, W., Yin, W., et al.: Feature extraction algorithm using a correlation coefficient combined with the VMD and Its application to the GPS and GRACE. *IEEE Access* (99), 1 (2021)
4. Sikka, G., Dhir, R.: Novel grey relational feature extraction algorithm for software fault-proneness using BBO (B-GRA). *Arabian J. Sci. Eng.* **45**(4), 2645–2662 (2020). <https://doi.org/10.1007/s13369-020-04445-2>
5. Wang, R., Shi, Y., Cao, W.: GA-SURF: a new speeded-up robust feature extraction algorithm for multispectral images based on geometric algebra. *Pattern Recogn. Lett.* **127**, 11–17 (2019)
6. Wang, N., et al. Feature extraction algorithm of a precession target based on a Doppler frequency profile of dual-aspect observation. *Appl. Opt.* **58**(17), 4695–4700 (2019)
7. Lu, T., Zhong, X., Zhong, L., et al.: A location-aware feature extraction algorithm for image recognition in mobile edge computing. *Math. Biosci. Eng.* **16**(6), 6672–6682 (2019)
8. Yin, D., Zhou, S., Wang, P., et al.: A feature points extraction algorithm based on adaptive information entropy. *IEEE Access* (99), 1 (2020)
9. Jiang, M., Yang, H.: Secure outsourcing algorithm of BTC feature extraction in cloud computing. *IEEE Access* (99), 1 (2020)
10. Attallah, B., Serir, A., Chahir, Y.: Feature extraction in palmprint recognition using spiral of moment skewness and kurtosis algorithm. *Pattern Anal. Appl.* **22**(3), 1197–1205 (2018). <https://doi.org/10.1007/s10044-018-0712-5>
11. Zhang, Y.: Image feature extraction algorithm in big data environment. *J. Intell. Fuzzy Syst.* **39**(1), 1–10 (2020)
12. Alharbi, A., Alghahtani, M.: Using genetic algorithm and ELM neural networks for feature extraction and classification of type 2-diabetes mellitus. *Appl. Artif. Intell.* **33**(1–4), 311–328 (2019)



Characteristics of Piano MOOC Learning Needs Based on Data Mining Technology

Yaokun Yang¹ (✉) and Allan Comstock²

¹ College of Music Liaoning Normal University, Dalian 116029, Liaoning, China
yyang@lnnu.edu.cn

² Emporia State University, Emporia, USA

Abstract. In recent years, the application of online courses is becoming more and more widespread, and the analysis and mining of education big data is becoming a new research direction. Based on the analysis of the characteristics of students' needs in the distance education of piano MOOC, this paper systematically studies the factors and rules of guiding students in the distance education, and tries to design and construct an effective distance learning mechanism for students, so as to realize the purpose of improving the teaching quality of distance education. This study aims to study the design of learning demand characteristics analysis of piano MOOC based on data mining algorithm. This document first introduces the theory of data mining. In addition, it also analyzes four parts of the learning needs of PIANO MOOC. Secondly, the model is designed based on the questionnaire survey of students, taking the questionnaire survey results of these students as a variable. Finally, the characteristics of learning demand characteristics are analyzed according to the actual situation. According to the data mining algorithm, the number of people using MOOC accounts for nearly 32% of the total, and the results show that the characteristics of learning needs can be divided into four dimensions.

Keywords: Data mining technology · Piano MOOC learning · Demand characteristics analysis

1 Introduction

In recent years, new knowledge production methods represented by online teaching and MOOC have become more and more popular among educational institutions and students [1]. At the same time, other technologies such as data mining and cloud computing are maturing [2]. In the past two years, due to the outbreak of COVID-19 at home and abroad, the concept and practice of MOOC has become more and more popular teaching mode at home and abroad [3]. MOOC are used by a variety of students, each with a very different learning purpose, age and knowledge background, their learning styles and interests are different [4]. Therefore, on the basis of mathematical statistics and data mining for piano MOOC learning demand characteristics analysis, through the analysis of a large number of data, we must provide effective data support analysis between the learning behavior

and demand characteristics in the survey report, and develop customized personalized learning functions [5].

With the development of data mining technology, massive open online courses are more and more widely used [1]. The discussion on MOOC has become a hot topic in recent years [6]. By analyzing the data set of MOOC, the learning behavior of users with the time characteristics extracted by Gao L and calculating the similarity of the time characteristics, k-means, SpectrClustering and Agglomerate clustering were used to analyze the experimental data. The experimental results under different cluster numbers and different data sizes were compared [7]. We use the silhouette coefficient to evaluate the validity of the clustering algorithm [8]. His research direction is very much in line with the current needs, but the experimental process is too complicated and practical [9, 10].

The innovation of this paper is to analyze the characteristics of students' learning needs according to the results of the questionnaire survey, which points to a new direction for data mining. The analysis results put forward some suggestions for the platform, schools and teachers to improve the level of piano MOOC education.

2 Data Mining Algorithms for Piano MOOC Learning Requirements

2.1 Data Mining Algorithm

Traditionally, data mining is defined as the process of extracting potentially useful knowledge and information from a large amount of incomplete, fuzzy and random data. It just means extracting some useful data from confusing information and data. Some units in China usually use data mining technology to deal with some businesses. The company analyzes a large amount of data by computer not only for research needs, but also to analyze and extract some valuable information from these data. The source of data mining methods is mainly the development of machine learning and artificial intelligence, combined with traditional statistics and imaging technology of scientific information technology, the formation of data mining technology. Data mining is the application of data information system and frontier defense. It is the development of complete database, expert system and network visualization in the field of interdisciplinary technology integration and various new fields.

2.2 CHAID Decision Tree Algorithm

There are many ways to obtain effective classifiers from existing large amounts of data, and decision trees are one of them. It is widely used in data mining, especially in data classification. Decision tree algorithm is a method to induce decision tree through a set of sample input data. The representation of a decision tree is an inverted tree structure diagram, and the probabilistic results are usually connected to the nodes of the branches. It is an intuitive graphical representation that uses statistical probability to analyze objects. After the decision tree model is established, it is necessary to give the evaluation value of the model to judge its advantages and disadvantages. Learning algorithm models

use training sets to build models so that models can be evaluated against calibration sets. The decision tree is evaluated after checking the total score. Evaluation indicators should include accuracy of classification, conciseness of description and complexity of calculation. CHAID decision tree algorithm is a common algorithm. To judge whether decision tree algorithm is used correctly, it depends on whether the difference between the two groups of decision tree grouping is large enough. The greater the difference in attributes, the more suitable the algorithm is.

CHAID decision tree provides a scheme to automatically search the variable most correlated with the dependent variable among various independent variables. The correlation measure adopts the chi-square value of significance degree, which can be calculated by formula 1 as follows:

$$\partial^2 = \frac{n(ab - cd)^2}{(a + c)(d + b)(a + d)(b + c)} \quad (1)$$

2.3 Piano MOOC Learning Needs

With the advent of the Internet era, the connotation of piano teaching has been enriched. For piano teaching there is no on-site practice, for students or want to cultivate the sentiment of adult learning, feeling is not so profound. Therefore, it is necessary to analyze online learning needs of piano MOOC. First of all, what students pay most attention to is practical ability. Students' learning needs tend to pursue practicality. We find that distance learning learners hope to get practical benefits from distance learning as soon as possible, such as job hunting and re-employment, and pay attention to the correlation between learning content and career development and personal value realization. The current education does not fully meet the actual needs of distance learning learners, indicating that there is a difference between the actual ability and the actual needs of students, which needs to be further improved. Coupled with the lack of face-to-face communication with teachers, many piano academic knowledge can not be flexibly applied in practice.

Secondly, the conflict between students' work and study time. In distance education courses, most students have to bear the dual pressure of career and study, so they can't have as much time to work as they study, and they face more learning difficulties than ordinary full-time students. Most distance learning students see a clear conflict between study time and work or other activities. How to arrange the time of study and MOOC reasonably is also a problem that the MOOC platform needs to adjust.

Finally, students' need for effective communication. Educational activity is a two-way movement between "teaching" and "learning" with knowledge and ability cultivation as the core. Distance education students need to communicate effectively with teachers in order to understand the key points of the piano. Among them, the structure diagram of the four parts of students' learning needs of PIANO MOOC is shown in Fig. 1:

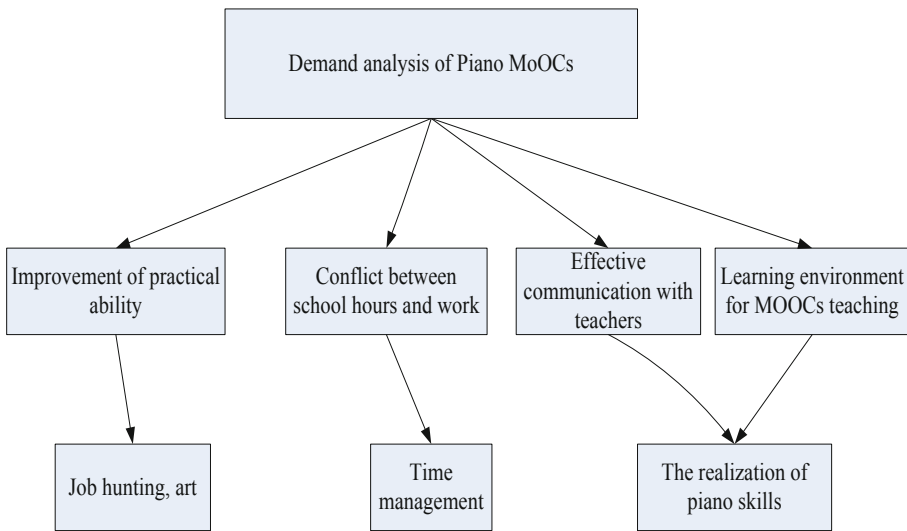


Fig. 1. Demand Analysis framework for Piano MOOC

3 Demand Analysis Framework for Piano MOOC

3.1 Model Design

This study focuses on assessing students’ attitudes towards piano MOOC and online learning through questionnaires and data screening. This paper will comprehensively evaluate students’ attitude towards course evaluation and learning behavior, as well as systematic and objective indicators such as learning time and duration. The first question to define the study was what were the behavioral patterns of students on the assessment platform, and what ideas were given to them to improve the assessment. What are the differences between students who study piano in MOOC and those who study offline, and what are the different reasons for their choice.

Through the specific arrangement and questionnaire survey of the PIANO MOOC course, the state of learning different types of behavior and different behavior states are repeatedly analyzed: the course lasts for 3 months, during which students have to complete online video training and unit tests, and then students have to complete the final exam online. In addition, students must attend a 4-h course and orientation teacher must prepare a final thesis. At the end of the course, a total of 1000 questionnaires were collected, and 937 valid questionnaires were collected. In this study, the k-means clustering model is constructed, and the sample survey results are used as variables to cluster the samples. The purpose is to further subdivide the samples according to the personality characteristics of each sample. After CART soft calculation, the decision tree algorithm is established as formula 2:

$$(1 - 3)FACTOR - 2(= 0.137[Mode : cluster - 3](21; 0.667) \tag{2}$$

3.2 Demand Analysis Experimental Data of Piano MOOC

According to the questionnaire of students, the most attractive reasons for the mooc course are the low price or even free teaching resources, the freedom of learning, the novel teaching methods and the teaching by famous teachers, which can almost meet the learning needs of students. When choosing the piano MOOC, students are more motivated by their personal interests and the idea of learning more piano playing skills. They hope to learn knowledge, master a knowledge and skills, expand their knowledge through the piano MOOC, and hope that the knowledge they have learned can be helpful to their future work and life. Through the course, students learned that knowledge contributes to real life, increased communication with teachers and students from all over the country, and a deeper understanding of their own professional courses and professional courses. At the same time, some learning needs are not met. For example, teachers sometimes fail to timely solve the problems of students' piano playing skills and rhythm, seldom participate in online discussions, and lack of learning atmosphere and motivation without studying together with students.

4 Demand Analysis of Piano MOOC Learning Based on Data Mining Algorithm

4.1 Status Quo of Online MOOC

Compared with other subjects, online courses have many advantages: it can provide students with cheaper learning opportunities, make the teaching environment richer teaching resources and so on. Network curriculum is the field of education and the discipline system, create a kind of advanced science and technology information technology development and the diversification of the network environment, the education system and education content improved, also greatly improve the teaching quality, the education system will be one of the most open and the development direction of the most dynamic, the teaching content is the most timely and targeted, widening the education on the network course platform can open The vision of students and teachers to improve the quality and level of scientific research.

In our country, many experts and scholars have been attaching great importance to the development and provision of online course construction and have done a lot of research and practice. In ensuring the concrete development of online course, the country has implemented financial and political support for online course construction, so that it can be easier to develop online course. Considering the results of online course design, which includes elementary, high school, college and other areas of online courses, the quality and quantity of its courses are rapidly increasing, and its scope is also expanding. However, for the study of piano, which is especially practical, the development of piano MOOC is developing rapidly, but it is not as fast as other courses. The data in Fig. 2 can be clearly compared with the development rate.

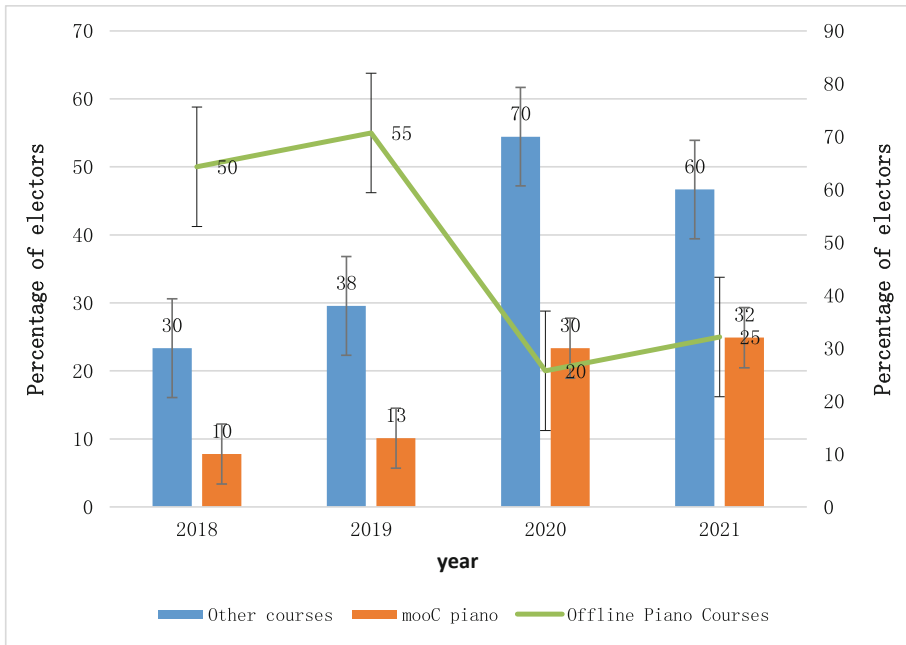


Fig. 2. Comparison of MOOC development from 2018 to 2021

4.2 Overall Demand for Piano MOOC

In this paper, the overall demand for online piano MOOC includes learning freedom, free course resources, innovative teaching methods and teaching by famous teachers. The overall demand analysis of online piano MOOC education is the data statistics and analysis of the characteristics of the four categories of overall demand.

Table 1. Characteristics of learning needs

Demand characteristics	The number of votes	Proportion (%)
Learning freedom	839	89.5
Free course resources	649	69.3
Innovative teaching methods	367	35.2
Famous teachers teaching	343	36.7

As can be seen from Table 1, among the 937 subjects interviewed, the number of people who want to learn freely and the demand is the highest, 839 people want to learn freely, accounting for 89.5%, followed by course resources free, accounting for 69.3%, and the demand for teaching by famous teachers, accounting for 36.7%, indicating that students want to learn more piano playing skills through MOOC. Be able to learn the

knowledge that you are interested in, master a knowledge and skill, and broaden your knowledge.

5 Conclusions

Based on the survey data, this paper selects 937 research samples and carries out data mining analysis on the characteristics of the samples. Based on the results of students' questionnaire survey, this paper provides areas for the school, MOOC platform and teachers to improve. As a remote mu originators of class education, should be equal treatment to each student, try our best to create a relatively fair for them and comfortable learning environment on the basis of analyzing the characteristics of the remote education student demand, proposed the research factors and laws of long-range teaching guide and motivate students outlook, strive to design and construct the effective distance learning students incentive mechanism, To improve the teaching quality of distance education.

References

1. Hendriks, R.A., Jong, P.G.M.D., Admiraal, W.F., Reinders, M.: Uncovering motivation and self-regulated learning skills in integrated medical mooc learning: a mixed methods research protocol. *BMJ Open* **10**, 38235 (2020)
2. Moustakas, A.: Spatio-temporal data mining in ecological and veterinary epidemiology. *Stoch. Env. Res. Risk Assess.* **31**(4), 829–834 (2017). <https://doi.org/10.1007/s00477-016-1374-8>
3. Li, J.Q., Yu, F.R., Deng, G., et al.: Industrial internet: a survey on the enabling technologies, applications, and challenges. *IEEE Commun. Surv. Tutor.* **19**(3), 1504–1526 (2017)
4. Ruipérez-Valiente, J.A., Halawa, S., Slama, R., Reich, J.: Using multi-platform learning analytics to compare regional and global MOOC learning in the Arab world. *Comput. Educ.* **146**, 103776 (2020)
5. Wu, S., Yang, J., Jin, C., Yu, X.: Anhui Sanlian University: Research on MOOC learning platform aided targeted teaching—a case study of Anhui Sanlian university. *China Comput. Commun.* (2019)
6. Rong, T., Min, X.: Analysis on higher-order discrete MPT model of MOOC learning behavior data. *Statist. Decis.* **10**, 14–17 (2019)
7. Gao, L., Zhao, Z., Qi, L., et al.: Modeling the effort and learning ability of students in MOOC. *IEEE Access* **7**(99), 1–12 (2019)
8. Chen, O., Woolcott, G., Sweller, J.: Using cognitive load theory to structure computer-based learning including MOOC. *J. Comput. Assist. Learn.* **33**(4), 293–305 (2017)
9. Krishnaiah, V., Narsimha, G., Subhash, N.: Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. *Int. J. Comput. Appl.* **136**(2), 43–51 (2016)
10. Wood, R.T., et al.: Problem gambling on the internet: implications for internet gambling policy in North America. *New Media Soc.* **9**(3), 520–542 (2016)



Application of Grey Model Algorithm in Urban GDP Growth Calculation Model System

Chenhe Wang^{1(✉)} and Arnold Salave²

¹ Major in Financial Mathematics, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China

² Jose Rizal University, 80 Shaw Blvd, Mandaluyong, Metro Manila, Philippines

Abstract. With the development of economic globalization, economic globalization has significantly promoted global economic growth, strengthened the economic ties between countries in the world, and improved the efficiency of global resource allocation. But on the other hand, economic globalization will also cause domestic economic fluctuations along with the economic impact of other countries, bring uncertainty to economic growth, and exacerbate the growth imbalance between developed and developing countries. This article takes GDP growth as an example, and inputs the signal generated by the gray model algorithm as a special excitation signal source into the PSPICE software. Taking Monte Carlo analysis as an example, the growth model PSPICE prediction results are imported into the gray model algorithm, and the data is post-processed to realize the data analysis between the PCPICE software and the gray model algorithm. Therefore, the combination of gray model algorithm and GDP growth for joint calculation can improve the accuracy of GDP growth. It can also post-process the forecast results to improve the accuracy of GDP simulation forecasts to a certain extent, thereby providing national GDP forecasts. The final research results show that through data analysis and numerical calculation of economic GDP growth, an economic growth prediction model has been established, the total GDP of a specific place from 2013 to 2020 has been studied, and the trend of local GDP has been predicted more accurately.

Keywords: Monte Carlo analysis · Grey model algorithm · Calculation model · GDP growth

1 Introduction

Economic globalization and the development of scientific and technological information technology have brought new opportunities and huge challenges to our country's economic development. In the modern information age, general programs for calculating the reliability of GDP growth are written using Matlab software. The reliability calculation program is divided into four parts: data input part, program calculation part (including random number creation part, reliability calculation part), output output and storage part for writing [1]. Based on the theoretical research of GDP forecasting model, this paper proposes a combination forecasting model with better application and better forecasting performance, which is of great significance to the research of GDP forecasting theory

and combination forecasting theory. This article adopts a new GDP forecast model in the actual forecast, which provides an effective method for the application of GDP forecast and opens up ideas for the choice of GDP forecast method.

In recent years, in the research of Matlab software economic GDP growth forecast model, many researchers have conducted research on it and achieved good results. For example, Gu Y believes that the economic development of a country's socio-economic structure covers the entire process of changes in material production resources. On the one hand, it is the accumulation of product quantity; on the other hand, it should also include improving product quality, improving production efficiency and increasing consumption [2]. From the perspective of harmonious development, Chanh K believes that the quality problems in the development process must be fully resolved. Forced growth is at the expense of consumption, ignoring the dynamic relationship between quantity and quality. The process of economic growth must be balanced and coordinated in the medium and long term, and the econoour of quality and quantity must go hand in hand [3]. At present, there are many researches on economic GDP growth forecasting modeling. These predecessors' theories and experimental results provide a theoretical basis for the research of this article.

This article analyzes the Matlab software, further demonstrates the far-reaching impact of GDP growth, and proposes a series of systematic studies on economic GDP growth, which pushes the Matlab software to a new height [4]. Matlab integrates many powerful functions in an easy-to-use window environment, such as arithmetic analysis, matrix calculation, scientific data display, and modeling and simulation of nonlinear dynamic systems. It is useful for scientific research, engineering design and people who need to perform arithmetic efficiently. Useful: many scientific fields provide complete solutions, and Matlab has convenient data display functions from the date of production.

2 Theoretical Basis of GDP Growth Calculation Model

2.1 The Theoretical Logic of the Quality of Economic Growth

(1) Economic growth maintains stable development

Economic growth must not only focus on technological contributions and progress, but also emphasize the sustained stability of economic growth and a coordinated and harmonious economic growth model. Among them, the stability of economic growth is reflected in two aspects: first, when economic fluctuations are caused by non-random factors in the economic system, the internal self-regulation and self-repair capabilities of the system are strong, and it has strong stability. The economic system can absorb this fluctuation in a relatively short period of time and return to its original stable state [5, 6]. The second is that when the economic system is unexpectedly impacted by the external econoour, it will break the initial steady state and produce a series of fluctuations caused by the impact, causing the internal economic function to face forced recession or false prosperity. If an econoour is stable, it refers to that the econoour has a strong ability to withstand external shocks when facing external shocks, and the effect of maintaining the stability of the internal operating state of the econoour is better.

(2) Inspection of the quality of macroeconomic growth

In the context of monitoring the quality of macroeconomic growth as the goal, the selection of monitoring and early warning methods often adopts the comprehensive prosperity index method. The comprehensive method of prosperity index monitoring and early warning is to divide the indicators according to the respective rate of change and time difference correlation rate of multiple indicators and the reference indicator period, divide them into the front group of indicators, and select a set of synchronization indicators with relatively complete and concentrated information. Among them, the selection of benchmark change indicators for early warning purposes is very important [7, 8]. The benchmark should be based on economic theory, that is, the choice should be based on economic theory. At the same time, the choice of benchmarks should truly and accurately reflect the changes and development directions of the early warning targets over the same period. When selecting indicators, one must choose a set of indicators that are real-time, sensitive, and can reflect the importance of the main economic growth quality. This is determined by the specific theoretical framework, derivative monitoring mechanism and early warning theory. The ultimate goal of monitoring and early warning is to provide regulatory guidance for policy adjustments based on the current dynamic timeline and possible macroeconomic quality trends in the future, which requires efficient indexed data selection.

(3) Theoretical support for economic growth forecast modeling

Based on the mechanism research and value judgment of the quality of economic growth in this article, the quality of a country's economic development is a complex system with comprehensive changes, characterized by nonlinear, unstable and unbalanced characteristics. Concerned about the coupling of economic development systems and social systems with natural resource systems. Therefore, the theoretical construction of the quality of macroeconomic growth includes many aspects and different dimensions [9]. The macroeconomic growth quality monitoring and early warning system is constructed from four aspects: the conditions of economic growth, the process of economic growth, the results of economic growth, and the benefits of economic growth.

2.2 Analysis of Predictive Control of Economic Growth

Forecasting includes 3 basic elements: forecasting model, rolling optimization and feedback correction.

- (1) The model used for prediction is called the prediction model. Many systems have the characteristics of nonlinearity, time variation, uncertainty and long time delay. It is very difficult to obtain an accurate mathematical model [10]. Predictive model, because the model can use previously known data to predict unknown information, it can be used as a predictive model instead of being stuck on the structure of the model, which greatly reduces the requirements for the model.
- (2) Predictive control is an optimized control algorithm that determines control variables through specific performance indicators. The optimization goal in prediction is not an offline execution, but an online repeated execution. At the sampling moment,

the optimized performance index only covers a limited range in the future. At the next moment, the range moves backward to achieve scrolling optimization.

- (3) Because there are many uncertain factors in the real system, the invariant model cannot accurately describe the real system. This requires additional prediction methods to make up for the lack of model prediction. Therefore, in each control cycle, the forecast will determine a set of future control effects by optimizing performance indicators. However, in order to avoid model distortion and environmental interference, all these control effects are not applicable, only the current control effects are applied, and the optimization is continued again [11]. MatlabMPCToolbox provides a series of functions for designing and analyzing predictive models. Using these characteristics, it is convenient to design GDP calculation program, and obtain intuitive system prediction results through GDP calculation program [12].

3 Research on Experimental Preparation of GDP Growth Calculation Model

3.1 Experimental Method

At present, prediction accuracy is used to measure the effectiveness of prediction models, which can avoid the influence of dimensions in different sequences. The prediction efficiency of the model is measured by the expected value of the prediction accuracy and the mean square error. In this work, efficiency indicators are used to look at individual forecasting models before combining forecasts.

- (1) Forecast validity index method

Let the time series be $\{x_t, t = 1, 2, \dots, n\}$, $\{x_{it}, t = 1, 2, \dots, n\}$ is the prediction value of the i -th prediction model at time t , if there is m single prediction models, namely $i = 1, 2, \dots, m$; $t = 1, 2, \dots, n$. Define the prediction relative error of the i -th prediction model at time as

$$\varepsilon_{it} = \frac{x_t - x_{it}}{x_t} \tag{1}$$

The prediction accuracy a_{it} of the i -th prediction model at time is as follows.

$$a_{it} = 1 - |\varepsilon_{it}| \text{ when } |\varepsilon_{it}| \leq 1 \tag{2}$$

- (2) Then the prediction validity $M f (f)$ of the i -th prediction model is defined as

$$E(A_i) = \frac{1}{n} \sum_{t=1}^n a_{it} \tag{3}$$

In the above formula, $I = 1, 2, \dots, M$, $e (a)$ is the expectation of the prediction accuracy sequence of the i th prediction model, which reflects the mean value of the prediction model at each time.

3.2 Experimental Data Collection

This article mainly uses Matlab2.0 software for data analysis and processing. Collected the GDP growth data of a certain city from 2013 to 2020, GDP growth rate = the current year's new gross product value/the previous year's gross product value. Generally speaking, with the development and progress of science and technology, people's living standards The increase in consumption, thereby increasing the level of consumption and income. The GDP will gradually rise, showing a positive growth trend. However, due to some special events or economic turmoil and other reasons, GDP may also experience negative growth. This article analyzes GDP growth to predict the city's economic development level.

4 Experimental Analysis of GDP Growth Calculation Model

4.1 Estimated Results of GDP Growth Model

According to the statistics of the total value of GDP in each year of 2013–2020 in a certain place, the relationship between GDP and standard deviation and P value is obtained, and the growth trend of GDP is analyzed according to the growth trend of GDP. The analysis data is shown in the figure below.

Table 1. Analysis results of GDP growth model for a certain place

Year	GDP (Trillion)	Standard error	P value
2013	0.2361	0.0612	0.0010
2014	0.3517	0.2131	0.1150
2015	0.4144	0.2104	0.0010
2016	0.6436	0.1180	0.0530
2017	0.8292	0.2410	0.6850
2018	0.7872	0.0846	0.3150
2019	0.8651	0.0114	0.0000
2020	0.9842	0.1524	0.0050

As shown in Table 1, from the analysis results of the GDP growth model for this region, it can be seen that the parameter estimation results of different time series under different regional states are mostly significant, and the regression parameters of each series are obviously different, indicating the two stages of construction in this article. The Markov band transition model can more accurately describe the dynamic change of economic growth in the zoning system. GDP has basically maintained a steady growth trend from 2013 to 2020. However, the growth trend from 2017 to 2018 shows a downward trend. Other than that, it is in an upward phase. Local GDP growth forecast has little impact. The results are shown in Fig. 1.

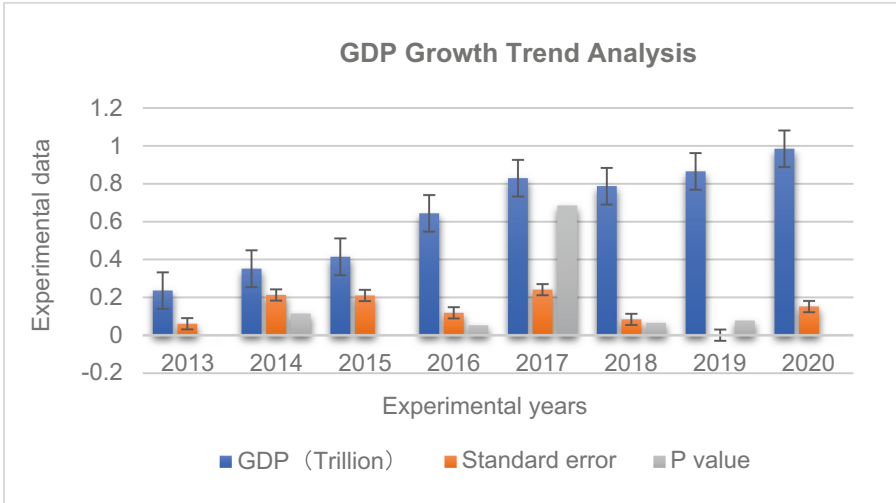


Fig. 1. Analysis results of a certain place’s GDP growth model

4.2 GDP Growth Trend Analysis

The GDP of this place has generally shown a growth trend in the past 8 years. The GDP growth rate analyzed in Fig. 2 is used to predict the approximate value of the total GDP of this place in 2021. The analysis is shown in the following figure.

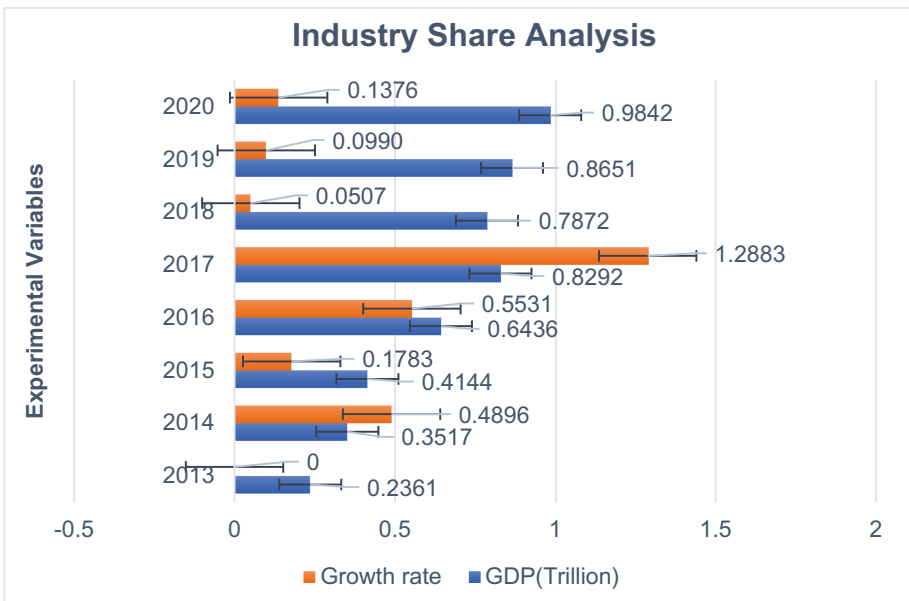


Fig. 2. Analysis of a certain place’s GDP growth trend

According to the data in Fig. 2. It can be seen that from 2013 to 2017, it maintained a good positive growth trend, and there was a slight decline in 2018, which may be due to the avian influenza virus at that time which caused the GDP of the poultry industry to decline, thus This has led to a decline in the local GDP. The GDP growth rate of 0.1376 in 2020 exceeds 0.0990 in 2019. It can be roughly predicted that the growth rate in 2021 will continue to rise, exceeding the ratio of 0.1376, and it is hoped that it will exceed 1.2 trillion. GDP total.

5 Conclusions

This article uses the powerful graphical editing function of Matlab software to provide graphical support for GDP growth trend analysis, and conducts visual and intuitive data analysis to facilitate the establishment of GDP growth models. This article focuses on the use of Matlab technology to link GDP growth models. This article mainly analyzes the GDP growth trend and explains the actual application of Matlab in detail to make the GDP growth forecast as close to the real value as possible. Finally, based on the characteristics of Matlab software forecasting that can improve forecasting accuracy, a method of applying forecast validity index method to study our country's GDP forecast is proposed. Aiming at the problem of less research on the projection of individual prediction models when constructing predictive effectiveness index method models, a complete model selection process is proposed. This article is based on the relevant knowledge of time series, taking GDP data of specific cities from 2013 to 2020 on the National Bureau of Statistics website as a sample. First, using all the data, with the help of Matlab software, research the development law of GDP data suitable for the site model. After a comparative analysis of the changes in annual GDP data, the predicted value is also close to the true value.

References

1. Zhu, C.: Research on mathematical modeling strategy based on MATLAB from the perspective of innovative ability training. *J. Phys. Conf. Ser.* **1648**, 022017 (2020)
2. Gu, Y., Chen, C.: Research on the model of regional circular economics based on computer modelling. *IOP Conf. Ser. Earth Environ. Sci.* **692**(4), 042064 (2021)
3. Chanh, K., Kushal, K., Vijay, T., et al.: On the growth of intensity forecast errors in the operational hurricane weather research and forecasting (HWRF) model. *Quart. J. Royal Meteorol. Soc.* **144**, 1803–1819 (2018)
4. Eldirdiry, O., Zaier, R., Al-Yahmedi, A., et al.: Modeling of a biped robot for investigating foot drop using MATLAB/Simulink. *Simulat. Model. Pract. Theory* **98**, 101972 (2019)
5. Wang, J.: Research on spatial conceptual modeling of natural language processing based on deep learning algorithms. *J. Phys. Conf. Ser.* **1345**, 042090 (2019)
6. Zhang, L.: Credit evaluation of medium and small sized enterprises during supply chain finance based on BP neural network. *Revista De La Facultad De Ingenieria* **32**(3), 776–784 (2017)
7. Deng, F.: Research on the applicability of weather forecast model—based on logistic regression and decision tree. *J. Phys. Conf. Ser.* **1678**, 012110 (2020)

8. Yang, J., Ma, L., Liu, J.: Modeling and application of ship density based on ship scale conversion and grid. *Ocean Eng.* **237**(10), 109557 (2021)
9. Qunxiong, Z., et al.: Research and application of causal network modeling based on process knowledge and modified transfer entropy - ScienceDirect. *IFAC-PapersOnLine* **51**(18):303–308 (2018)
10. Sayari, N., Sari, R., De Hammou, H.S.: The impact of value added components of GDP and FDI on economic freedom in Europe. *Econ. Syst.* **42**(2), 282–294 (2018)
11. Zkan, A.N.: Modeling of energy consumption forecast with economic indicators using particle swarm optimization and genetic algorithm: an application in Turkey between 1979 and 2050. *Alphanum. J.* **8**, 59–78 (2020)
12. Nedi, V., Despotovi, D., Cvetanovi, S., et al.: Institutional reforms for economic growth in the Western Balkan countries. *J. Pol. Model.* **42**(5), 933–952 (2020)



Design and Implementation of Nursing Big Data Platform Based on Data Mining Technology

Dapeng Liu¹(✉), Hui Zheng¹, Qiongqiong Yan², Ying Li³, and Cuicui Huang¹

¹ Zaozhuang Vocational College of Science and Technology, Tengzhou 277599, Shandong, China

² Tengzhou Central People's Hospital, Tengzhou 277599, Shandong, China

³ Zaozhuang Cancer Hospital, Tengzhou 277599, Shandong, China

Abstract. With the development of society, people's living standards have been greatly improved, and there are higher requirements for medical services. Nursing big data platform is a new efficient, convenient and personalized nursing technology platform. This paper designs a data mining technology based on data mining technology, combined with the amount of nursing big data for mining and analysis, and puts forward a scheme to improve the existing problems in hospital nursing and the impact of solving these problems. Finally, the performance of each module of nursing big data platform is tested. The test results show that the average response speed is basically about 5.0 m/s. The delay time is about 2S, accounting for about 3100K of memory, and accounting for about 2% of the CPU of the whole device. This shows that the performance of the nursing big data platform designed in this paper basically meets the needs of users.

Keywords: Data mining · Nursing technology · Big data · Platform design

1 Introduction

With the development of network technology, the traditional nursing big data platform can no longer meet the information needs of users. People put forward higher requirements for the quality of life and nursing level, especially in medical treatment, which brings new opportunities to the hospital [1, 2]. Therefore, it is urgent to use data mining, machine learning and other methods to solve practical problems [3, 4].

Many scholars have done relevant research on data mining technology. Based on medical Internet big data, some scholars have completed the development and design of the analysis platform based on internet medical big data by using visualization, data mining and other technologies, and realized the functions of hospital importance related index analysis, regional expert doctor recommendation, regional key hospital ranking and so on [5, 6]. There are also scholars' basic methods of analyzing medical big data based on data mining technology, and give specific application examples [7, 8]. The above research has laid a certain foundation for the research of nursing Digitization Based on data mining technology.

This paper first introduces the design process of nursing academic platform, and then analyzes the nursing academic architecture, functional principle and database model. Finally, through practical examples to verify its feasibility, availability, security and other performance indicators and result stability analysis, the results show that nursing big data mining technology can provide more efficient and convenient services for the majority of users.

2 Discussion on Nursing Big Data Platform Based on Data Mining Technology

2.1 Nursing Data

Nursing computerization is a part of hospital computerization, and there is no clear concept at present. Some only define it as the use of modern network technology, computer technology, communication technology, etc. The traditional nursing work mode should constantly respond to the improvement of the overall level of nursing and speed up the process of modernization [9, 10].

At present, nursing big data is mainly reflected in nursing big data system, which is an important part of hospital information system. It mainly includes two platforms: nursing management and clinical nursing. It is divided into mobile monitoring system, intensive care system, electronic medical record system, anesthesia monitoring system, shift planning system, personnel management system and other subsystems. Its main functions include clinical nursing, medical prescription management, medication management, disease monitoring, project management, appointment management, quality control, etc. [11, 12].

Nursing big data is to use Internet of things, big data, cloud computing and other technologies to intelligently meet users' diversified daily medical needs and provide users with convenient and fast medical services. Medical and health will build intelligent and logical links of medical information and build a big data-driven medical ecosystem. With the big data of nursing, users can not only meet their medical needs, but also benefit from preventive health care services. Medical and health services are a complex group, including disease prevention, health care, disease diagnosis and treatment, surgery, rehabilitation, scientific research and so on. Intelligent medical intervention has brought great opportunities and challenges to medicine and medical field. Intelligence is the most important feature of medical big data.

2.2 Impact of Big Data on Nursing Industry

With the rapid development of the information internet, the importance of big data technology has been paid more and more attention by the whole industry. Wikipedia points out that big data is a large amount of complex data that traditional data processing methods can not handle. Big data can accurately analyze user behavior, provide basis for decision-making, and carry out targeted business activities. The integration of big data and nursing industry can help nurses make better use of their professional knowledge,

skills and learning, so as to reduce nursing links and finally improve nursing quality. With the continuous acceleration of computerization, the medical and health field has gradually entered the era of big data. Big data has great potential to improve health and solve various complex problems, such as infectious disease prevention and management, collaboration, people-oriented care and the prevention of chronic diseases through personal lifestyle and knowledge.

2.3 Data Mining Technology

2.3.1 Definitions

There are many contents in data mining because it involves a wide range of disciplines and is developed based on statistics, applied mathematics, machine learning and other disciplines. In the broadest sense, the knowledge that can be extracted from data can be called the content of data mining. Data mining has the following three characteristics:

Firstly, the amount of data in data mining is often huge. Therefore, how to access data effectively and how to determine the relationship between data according to specific application fields. Efficient algorithms to find and use all data or random or intentionally selected data subsets become very important for data mining workers.

Secondly, the data generated in data mining is often collected for other purposes. When collecting data, one or more important variables may not be collected. These variables will be useful and even important for exploring data in the future.

Third, another feature of data mining is that data miners are usually reluctant to include a priori knowledge into the algorithm, because it is equivalent to performing hypothesis testing (but statistical hypothesis testing is not excluded as part of it).

2.3.2 Data Mining Process

Figure 1 is the basic process of data mining.

Data mining process model provides macro instructions and engineering methods for data mining. A reasonable processing model can organically combine various processing steps and guide users to better develop and use data mining systems. Figure 1 shows the data mining process from a technical point of view. It focuses on optimizing data preprocessing and omits many complex steps required in the nursing industry. This data mining view shows how to export useful models from raw data.

2.3.3 Data Mining Algorithm

For data record D , $I = \{i_1, i_2, i_3, \dots, i_m\}$ is a set of M different elements. Each transaction D is composed of these elements, that is, $D \in I$. The mined association rules have the form of $X \rightarrow Y$. The $X \rightarrow y$ rules are established only when the minimum support threshold minsup and the minimum trust threshold minconf are reached. The probability of the rule is represented by $P(XUY)$ and $P(Y | X)$. Support is the probability that the set of elements a and B will appear when they contract, and confidence is the probability that the set of element X will also appear in the transaction in which element Y appears. The

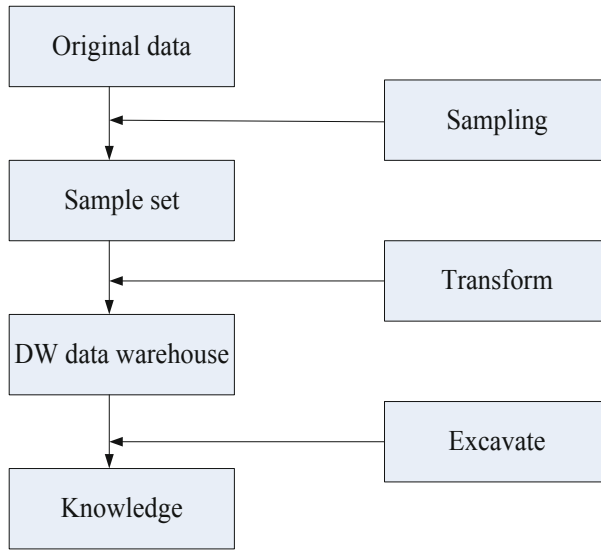


Fig. 1. The process of data mining

problem of mining association rules is to find the minimum minsup support and minconf trusted association rules set by the user in transaction database D . Association rule $X \rightarrow Y$ has a degree of support for sup, that is, sup is the percentage of X and Y transactions in D , that is, probability $P(XUY)$, that is:

$$\text{Sup}(X \Rightarrow Y) = \text{Sup}(XUY) = P(D) \quad (1)$$

Rule $X \rightarrow Y$ has credibility C in the database, which means that C is the probability of including both itemset A and itemset B , which is the conditional probability $P(Y|X)$, that is:

$$\text{Conf}(X \Rightarrow Y) = \text{Sup}(XUY) / \text{Sup}(x) = P(Y|X) = P(XY)P(X) \quad (2)$$

The research of association rules can be divided into the following two processes:

- (1) Find all common element collections: the occurrences of these element collections must be greater than or equal to the predefined minimum occurrences.
- (2) Obtain association rules based on common itemsets: these rules must also meet the minimum support and trust set.

3 Experiment

3.1 Nursing Big Data Platform Module

3.1.1 Mobile Nursing Staff Information

Collection and calculation functions (including access volume, body mass index, infusion speed, due date, etc.), doctor's order query, execution and statistics functions, bar code reading, consumables input and loading, test result query, etc.

3.1.2 Nursing Workload Statistics Module

The nursing stress statistics reporting system can use various index generation algorithms to analyze and calculate directly from the his system database, and automatically generate various nursing stress statistics, so as to reduce the workload of clinical nurses and nursing managers and achieve better results. The 254th Hospital of PLA used PDA to record every nursing operation, and then calculated the workload according to the channel, which improved the accuracy and speed.

3.1.3 Nursing Adverse Event Reporting Module

The reporting system of nursing adverse events is an important method and means of nursing safety management. The reporting procedures vary from mandatory, sometimes mandatory to voluntary, with their own advantages and disadvantages.

3.1.4 Nursing Product Management Module

The article management system fully discloses all links such as procurement, warehousing, outbound, inventory, quantity and price. The required quantity and specifications of various articles are clear at a glance to realize the efficient management of articles in the whole hospital. Some hospitals have also introduced bar code technology to improve the accuracy of management.

3.1.5 Nursing Staff Management Module

Nursing human resource management system realizes personnel recruitment information management, nurse learning, training, assessment, performance management and comprehensive management, so as to reduce labor intensity and improve management efficiency. Performance appraisal is an important part of human resource management. The computerization of performance appraisal can improve the management level and make it more practical, faster, more objective and more real.

3.2 Data Architecture

The data architecture is responsible for organizing data to meet various data requirements of the application system. It is an important part of building a flexible and powerful big data platform and the basis of analysis.

(1) Temporary data area

Temporary data area is the part used to load data. It is usually composed of data files and database objects. The definition of buffer layer should be coordinated with upstream and downstream areas. The data in the temporary data area is generally temporary and is only for system use, not for company employees to query.

(2) Publish source data area

The back source data area is a data area stored in the population data analysis platform, which is very consistent with the upstream source system and manages history. The data collected from the source system is in the post source data area, but not necessarily in the main data area. The post source data layer generally has three purposes: one is to manage the data from the source system for audit and monitoring requirements. Second, processing simple reports to improve application speed. Third, keep the original data to facilitate data verification. Fourth, retain the original data to facilitate later data integration.

(3) Feedback data area

The feedback data area is located between the release source data and the application area and serves as a bridge for the big data platform. Based on the business perspective, the feedback data area extracts the common data access and statistical requirements of the unified data platform, creates public data supporting applications, and provides services for accessing public data that can be processed multiple times at the same time. Organize data according to business analysis topics, conduct data integration and denormalization of specific granularity, and save historical snapshot data in important time.

(4) Application domain

The application domain is the main component of the external service of the unified data management platform. The application domain shall process the data of base domain, source domain, label domain and domain according to the requirements of their respective applications, and provide it to users or external applications. The application area includes application computing area and application access area. The purpose of the application computing area is to use the powerful computing power of the big data platform to calculate the consistency of data. The application access area is used to enable access to applications. Because the access speed of Hadoop platform is slow and can not respond to the functions of the application, the calculation results of the application area are synchronized with the relational database to facilitate the access of the application.

(5) Unstructured data area

A large amount of unstructured data carried by unstructured database is efficiently managed and centrally managed through big data to meet the real-time full-text retrieval needs of coupons, web pages, licenses, documents and other documents.

(6) Historical data area

It is used for long-term storage of historical data to realize parallel computing and real-time query of massive data. In addition to storing historical snapshot information of data, it also stores associated data, knowledge graph data, etc.

(7) Index area

An index of information used to store unstructured data, where the information is generated in a reverse structure. The so-called inverted line is to change the original structure of Document → Word into inverted word structure Word → Document. The advantage of inverted structure is that it can quickly create documents after localized word search, which is the basic work of full-text recovery.

4 Discussion

4.1 Performance Test of Nursing Big Data Platform

Table 1 shows all aspects of performance test data of nursing big data platform.

Table 1. Performance testing

Number of tests	Reaction speed (m/s)	Delay time (s)	Memory footprint (k)	The CPU occupancy rate is (%)
1	5.3	3	3125	2
2	6.3	2	3214	1.5
3	3.5	1	3253	2
4	4.3	3	2314	3
5	4.2	4	3625	2
6	4.6	5	3102	2
7	5.3	2	3410	3
8	4.1	6	3250	2
9	4.0	3	3142	3
10	4.5	2	3152	2

As can be seen from Fig. 2, in the performance test of the nursing big data platform, all modules were tested 10 times, the average response speed was basically about 5.0 m/s, while the delay time was basically about 2S, accounting for about 3100K of memory, and accounting for about 2% of the CPU of the whole equipment. This shows that the performance of the nursing big data platform designed in this paper basically meets the needs of users.

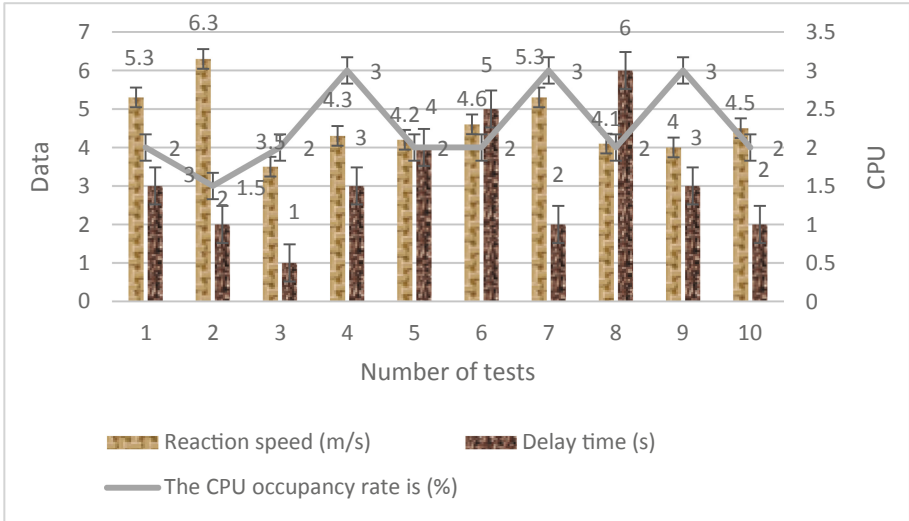


Fig. 2. Performance test comparison diagram

5 Conclusion

With the advent of the era of big data, nursing service has become a huge and valuable field. This paper mainly studies and analyzes the problems existing in the application of nursing technology in the large platform. Firstly, the database system based on data mining mode is introduced. Then, according to the model, medical information management is combined with other modules to realize its functions and performance indicators. Finally, a design method based on background management system architecture and related algorithm implementation scheme are proposed to solve the above problems, in order to provide users with better and faster service experience and meet their nursing needs.

Acknowledgements. Research and Practice on the Cultivation of Innovative Teaching Teams of “Three Types” Teachers Based on Mozi’s Vocational Education Thoughts (No. LX2021194).

References

1. Alameen, A.: Repeated attribute optimization for big data encryption. *Comput. Syst. Sci. Eng.* **40**(1), 53–64 (2022)
2. Annathurai, K.S., Angamuthu, T.: Sorensen-dice similarity indexing based weighted iterative clustering for big data analytics. *Int. Arab J. Inf. Technol.* **19**(1): 11–22 (2022)
3. Dubuc, T., Stahl, F.T., Roesch, E.B.: Mapping the big data landscape: technologies, platforms and paradigms for real-time analytics of data streams. *IEEE Access* **9**, 15351–15374 (2021)
4. Liu, K., et al.: Granular cabin: an efficient solution to neighborhood learning in big data. *Inf. Sci.* **583**, 189–201 (2022)

5. Shahbazian, R.: Distributed optimization over large-scale systems for big data analytics. *4OR* **19**(2), 309–310 (2021)
6. Alfred, R., Obit, J.H., Chin, C.P.-Y., Haviluddin, H., Lim, Y.: Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks. *IEEE Access* **9**, 50358–50380 (2021)
7. Awajan, I., Mohamad, M., Al-Quran, A.: Sentiment analysis technique and neutrosophic set theory for mining and ranking big data from online reviews. *IEEE Access* **9**, 47338–47353 (2021)
8. Shravani, D.D.: Novel design and implementation of graph mining for big data network analysis. *Int. J. Eng. Res. Appl.* **07**(7), 16–25 (2017)
9. Bhat, S.A., Huang, N.-F.: Big data and AI revolution in precision agriculture: survey and challenges. *IEEE Access* **9**, 110209–110222 (2021)
10. Li, Y.: Design and implementation of intelligent travel recommendation system based on internet of things. *Ingenierie des Systemes d'Information* **23**(5), 159–173 (2018)
11. Castro, A., Villagrà, V.A., García, P., Rivera, D., Toledo, D.: An ontological-based model to data governance for big data. *IEEE Access* **9**, 109943–109959 (2021)
12. Derhab, A., Alawwad, R., Dehwah, K., Tariq, N., Khan, F.A., Al-Muhtadi, J.: Tweet-based bot detection using big data analytics. *IEEE Access* **9**, 65988–66005 (2021)



Construction of the Intelligent Online Tourism English Model Corpus Based on ESP

Jianzhou Cui(✉)

Wuxi City College of Vocational Technology, Wuxi 214153, Jiangsu, China

Abstract. With the continuous improvement of social informatization and the increasingly fierce competition in the tourism industry, tourism English services have become an important means of communication. Online corpus is formed and established on the basis of language and text, which can realize the sharing of language information. With the vigorous development of the tourism industry, the service function of tourism English has become increasingly prominent. In order to meet the needs of social development for high-quality talents, China has continuously reformed its education. Therefore, domestic colleges and universities have set up courses such as online learning and remote access. At the same time, intelligence has become the target and direction of one of the research hotspots in the future information age. In this context, how to make better use of modern information technology to build a corpus of intelligent online tourism English models has become a topic worthy of discussion. This paper adopts questionnaire survey method and data analysis method to study the establishment of an efficient, practical and compatible intelligent online travel English model corpus based on ESP technology, and at the same time better meet the user's personality demand of the corpus resources. According to the survey results, most of the interviewees believe that the construction of the intelligent online travel English model corpus is necessary, and it occupies an important position in the development of translation work and other fields.

Keywords: ESP corpus · Corpus construction · Tourism English · Intelligent model

1 Introduction

In today's society, people are increasingly developing and utilizing information resources. With the increasing market demand of the tourism industry and the increasing pressure of competition in the tourism industry, many domestic enterprises have launched intelligent construction projects to improve service quality and efficiency and gain competitive advantages. The traditional tourism industry can no longer satisfy consumers' demand for high-quality and diversified consumption. The construction of an intelligent online English corpus based on ESP can solve the above problems well. Based on ESP technology, this article explores the construction of intelligent online corpus, which can enable users to obtain the required information resources and reference objects more

quickly, accurately and quickly, and can effectively help solve the existing problems of current intelligent services.

From the current point of view, the research results on the construction of ESP corpus and English corpus in academia are very rich. For example, Liu Ping proposed to conduct teaching experiments for academic English vocabulary learning based on the ESP corpus, which proved that the corpus is very effective in academic vocabulary learning [1]. Wang Wenhua believes that by studying the status quo of ESP and corpora at home and abroad, we have discovered the deficiencies of existing corpora in teaching. Therefore, the construction of higher vocational English corpus and sharing platform based on ESP is very necessary [2]. Yao Yao pointed out that the effective integration of professional English vocabulary teaching and corpus technology enhances students' confidence in vocabulary use in writing, and also improves the initiative of vocabulary learning [3]. Therefore, this article is based on ESP-based intelligent online tourism English model corpus construction research, which caters to the actual needs of English corpus construction, has certain practical significance and research value.

This article mainly explores from these aspects. First, it explains what ESP theory is and expounds its research status. Then it discusses the construction of the intelligent online tourism English model corpus. In addition, the application of ant colony optimization algorithm in the construction of intelligent online travel English model corpus is also introduced. Finally, a questionnaire survey is conducted on whether it is necessary to build a corpus of intelligent online tourism English model, and the survey results and analysis are obtained.

2 Research on Corpus Construction of Intelligent Online Tourism English Model Based on ESP

2.1 ESP Theory and Research Status

ESP, or English for Specific Purposes, began in the 1960s. It refers to English related to a specific job, profession, or purpose, which is the so-called English for Specific Purposes. As a new subject field, its emergence is an inevitable social and historical development, and it has many reasons for it.

ESP theory is based on the learner-centered teaching thought, and promotes the development of teaching content and methods and learners' learning goals and needs in a harmonious and unified direction. It also emphasizes the authenticity of the content of the teaching materials, the importance of the needs of learners, and the centrality of students. Therefore, this theory has systematic and scientific guidance for the use of English for specialized language teaching.

Many scholars agree that ESP should start from the needs of learners, and that ESP should be related to specific occupations. At the same time, we can see the importance of needs analysis in ESP teaching. Tourism English is produced in order to meet the requirements of tourism development for the English level of tourism practitioners. Therefore, understanding the needs of society and students is an indispensable link in improving the level of tourism English teaching [4, 5]. The ESP is divided according to the learning process of the learner, and the specific classification is shown in Fig. 1.

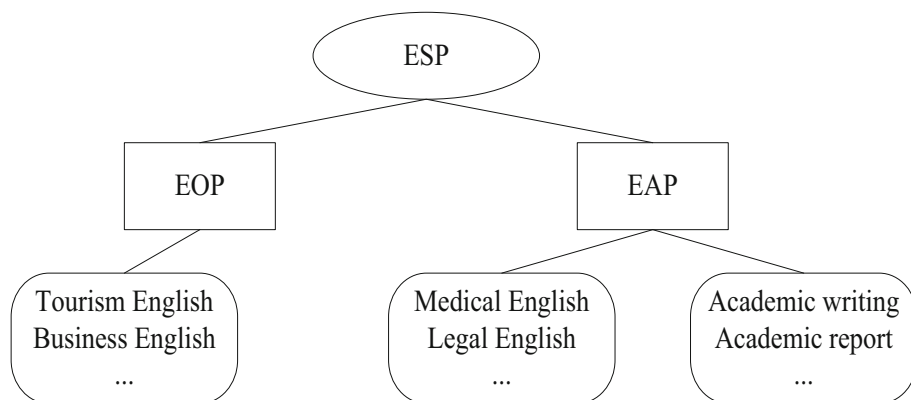


Fig. 1. Classification of ESP

Through the various efforts of domestic and foreign scholars and experts, ESP has achieved fruitful results in theoretical research and teaching practice, which has greatly promoted the development of ESP, and ESP has become more and more popular. In many English-speaking countries and non-English-speaking countries, various educational institutions have gradually formed a relatively complete ESP teaching and research system, which not only provides ESP courses, but also focuses on ESP theoretical research and English teaching, and awards corresponding diplomas [6, 7].

2.2 Construction of the English Model Corpus of Smart Online Tourism

The intelligent online travel English model, through the retrieval of relevant keywords in the corpus, in the process of analysis and research, the language information required by language learners is converted into recognizable sentence content, and converted into readable, understandable expression to express Chinese with high accuracy. At the same time, you can use the network platform to communicate with language learners online. In addition, it also meets the user's online translation and other functional requirements.

The corpus refers to the collection, sorting and classification of travel English information required by users according to certain standards. Corpus construction is the core content of Tourism English. It extracts textual information of items with the same characteristics from the database according to the language needs of different users, and on this basis, integrates and analyzes all data sets. The construction of the corpus is an important link in the direction of intelligent development of tourism English. It mainly includes the collection, sorting and storage of existing data, and thus the establishment of related databases [8, 9].

The ESP-based intelligent online travel English model corpus is mainly divided into three parts, including these parts. The first part is the database storage layer. The second part is the process of data collection and analysis, involving basic knowledge of related field classification and statistics. The third part is the keyword search module, data storage, online retrieval and automatic update. Data storage means that users can query the

corresponding content in the required file by entering relevant information; online translation includes operation processes such as text format conversion; automatic publishing means that the document is sent to the database server after the system is completed, and then Real-time tracking on this platform and feedback to the corresponding personnel.

The establishment of an intelligent tourism English corpus is a complicated process. In the construction of the intelligent tourism English corpus, all items should be classified, and then the content should be marked according to the functions that need to be provided under different categories. The following methods can be used in the translation process. First, the semantic method, since there is a certain relationship between the user input information and the output information or has some similar attributes, a database table and field set is constructed. Second, keyword retrieval, that is, analyze and organize the corresponding keywords in the query table, and then store them in the corpus [10, 11].

On this basis, the user can query the desired content by entering relevant information. The corpus then automatically extracts and online translation based on the information entered by the user, and then sends the document to the database server after completion, and finally converts the original text into a format that meets the requirements and can be used for users to read or apply.

2.3 Application of Ant Colony Optimization Algorithm in the Construction of Intelligent Online Tourism English Model Corpus

The ant colony optimization algorithm simulates finding the path of ants. In pathfinding, ants leave pheromone on the path, and the pheromone generates positive feedback under certain conditions, accelerating the formation of the optimal path. The ant colony algorithm uses this principle to create positive feedback on the global level from individual local information to speed up the optimization process. The ant colony optimization algorithm is also a part of the intelligent optimization algorithm. By imitating natural and biological phenomena, these optimization mechanisms are applied to optimization calculations, and a new algorithm optimization is developed. This algorithm is derived from random search, but it also has the ability to adapt to the environment [12, 13].

The intelligent embodiment of the algorithm has a certain degree of adaptability, which can continuously adjust the search strategy according to the algorithm during the search process to improve the search effect. Create different search strategies according to different simulation situations. The information collection model of ant colony algorithm is mainly a strategy to modify pheromone. With the development of ant colony algorithm, the strategy of modifying pheromone has become more diversified. The ant colony algorithm also has the intelligent characteristics of the intelligent optimization algorithm, which helps the intelligent online travel English model corpus to improve the search function section, continuously optimize, and promote the further development of the corpus construction work. The specific calculation formula is as follows.

$$\psi_{ug}^{l+2} = \psi_{ug}^l \cdot \varpi + A/M \quad (1)$$

$$O(y) = \prod_{u=2}^b O_{y_u} \quad (2)$$

Among them, M is the path, ψ_{ug}^l is the pheromone on the edge ug of the first generation. ω is the evaporation coefficient, which means that the pheromone will decrease over time. A is a constant used to control the amount of pheromone change. y_u is the u -th component of y .

3 Questionnaire Survey on the Construction of the Corpus of the Intelligent Online Tourism English Model

3.1 Questionnaire Design Process

The subjects of the questionnaire survey are 120 citizens of city B. Through the issuance of online questionnaires or paper questionnaires, the collection and quantitative analysis of the information filled in by users are carried out to draw conclusions of the questionnaire.

- (1) In the preliminary preparation of the questionnaire, the number of questions should be as concise as possible to avoid fatigue of the interviewees.
- (2) Questionnaire delivery The questionnaire is distributed through online questionnaires, on-site questionnaires, and inviting friends to help ask friends and students around them. A total of 120 questionnaires are distributed, 120 valid questionnaires are returned, and the questionnaire recovery rate is 100%.
- (3) Questionnaire analysis The collected questionnaire information will be sorted out and the required information data will be obtained. Analyze the results of the questionnaire. The analysis results include citizens' opinions on whether it is necessary to construct a corpus of intelligent online tourism English models. Some of the results obtained from the questionnaire are as follows.

3.2 Questionnaire Survey Content

The first part is the selection of 120 citizens of city B. The interviewees are divided into teachers, students, translators and travel enthusiasts according to their work and hobbies, and the necessity of constructing an intelligent online tourism English model corpus is investigated.

The second part is to sort out the information collected in the questionnaire and understand the attitudes of 120 citizens towards the construction of the intelligent online travel English model corpus. Part of the questionnaire survey results are as follows.

4 Questionnaire Survey Analysis on the Construction of the Intelligent Online Tourism English Model Corpus

This questionnaire survey divided the interviewed citizens into teachers, students, translators and travel enthusiasts according to their work and hobbies. A survey was conducted on whether it is necessary to build an intelligent online travel English model corpus. The results of the survey are shown in Fig. 2 (Table 1).

Table 1. Necessity of the construction of intelligent online tourism English model corpus

Project	Teacher	Student	Translator	Travel enthusiast
Very necessary	10	9	12	8
Necessary	12	10	10	11
General	7	8	6	7
Unnecessary	1	3	2	4

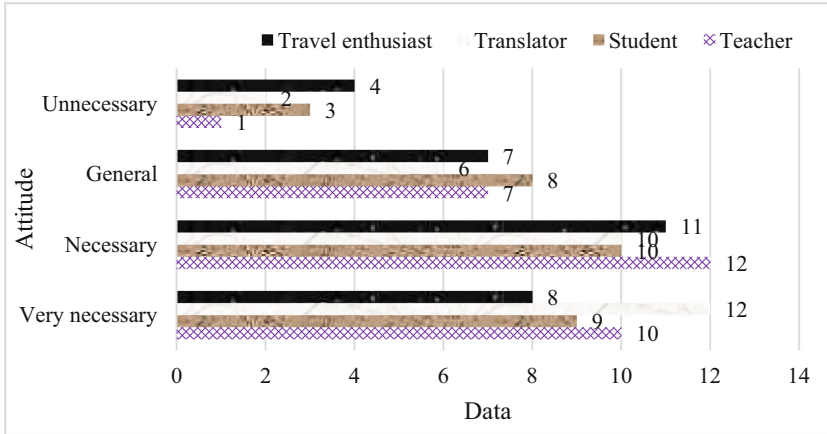


Fig. 2. Necessity of the construction of intelligent online tourism English model corpus

It can be seen from Fig. 2 that among the 120 interviewees, a total of 82 people indicated that it is necessary to build a corpus, while only 10 people indicated that it is unnecessary. Among them, 22 of the translators expressed the need for corpus construction. It can be seen that most of the interviewed citizens believe that the construction of the intelligent online travel English model corpus is necessary, and it occupies an important position in translation and other fields.

5 Conclusion

In recent years, the tourism industry has developed rapidly, and tourism English has become the focus of attention. With the popularization of Internet technology, tourism English, as a new type of service language, has more and more applications, involving all aspects of people’s daily lives. The construction of online corpus is to better improve the service quality of online platforms. This article summarizes the application status of the online travel English corpus and related materials. In addition, a questionnaire survey was launched and the conclusion was reached: it is very necessary to construct a corpus of intelligent online tourism English model. Therefore, this article builds an intelligent online tourism English model based on the ESP corpus, studies the process

of corpus construction, and is of great significance to the development and construction of English translation work.

References

1. Liu, P., Liu, Z.: Research on the effectiveness of academic English vocabulary learning method based on ESP corpus. *Foreign Lang. Res.* **35**(169(03)), 56–62 (2018)
2. Wang, W., Ma, Y.: Analysis on the construction of higher vocational English corpus and sharing platform based on ESP (2016-1), 142–144 (2021)
3. Yao, Y., Cheng, H.: The construction of professional English vocabulary teaching mode assisted by ESP corpus—taking tourism English classroom teaching as an example. *Educ. Mod.* **005**(049), 272–274, 317 (2018)
4. Andelmin, J.: Models and algorithms for vehicle routing, resource allocation, and multi-stage decision-making under uncertainty; Malleja ja algoritmeja ajoneuvon reititykseen, resurssien allokointiin, sekä monijaksoiseen päätöksentekoon epävarmuuden alla. Aalto University, Espoo, Finland (2021)
5. Kashyap, N.: Algorithms for Power System State Estimation with Phasor Measurement Units. Aalto University, Espoo, Finland (2021)
6. Choffin, B.: Algorithmes d'espacement adaptatif de l'apprentissage pour l'optimisation de la maîtrise à long terme de composantes de connaissance. (Adaptive spacing algorithms for optimizing long-term mastery of knowledge components). University of Paris-Saclay, France (2021)
7. Esteve, M., Rodríguez-Sala, J.J., López-Espín, J.-J., Aparicio, J.: Heuristic and backtracking algorithms for improving the performance of efficiency analysis trees. *IEEE Access* **9**, 17421–17428 (2021)
8. Mass-Sanchez, J., Vargas Rosales, C., Ruiz-Ibarra, E., Espinoza-Ruiz, A., Vázquez Castillo, J., García-Berumen, A.: Weighted multilateration algorithm with a correcting factor for mobile wireless sensor networks. *IEEE Access* **9**, 161937–161958 (2021)
9. Hernández-Sanjaiime, R., González, M., Peñalver Benavent, A., López-Espín, J.-J.: Estimating simultaneous equation models through an entropy-based incremental variational bayes learning algorithm. *Entropy* **23**(4), 384 (2021)
10. Feng, M., Yang, Y., Shu, Q., Yang, R.: An Improved ESPRIT-based algorithm for monostatic FDA-MIMO radar with linear or nonlinear frequency increments. *IEEE Commun. Lett.* **25**(7), 2375–2379 (2021)
11. Mazzei, A., Cerrato, M., Esposito, R., Basile, V.: Ranking algorithms for word ordering in surface realization. *Inf.* **12**(8), 337 (2021)
12. Botta, M., Cavagnino, D., Esposito, R.: NeuNAC: a novel fragile watermarking algorithm for integrity protection of neural networks. *Inf. Sci.* **576**, 228–241 (2021)
13. Bracino, A.A., et al.: Optimization of biofilter size for aquaponics using genetic algorithm. *J. Adv. Comput. Intell. Intell. Inf.* **25**(5), 632–638 (2021)



Data Mining of E-Commerce Enterprise User Preferences in the Context of Big Data

Lielong Yang¹(✉) and Xingtan Yin²

¹ Hechi University, Hechi 546300, Guangxi, China

² Guilin Normal College, Guilin 541199, Guangxi, China

Abstract. In the era of big data (BD), consumer behavior information is easier to collect. Therefore, as a popular e-commerce (EC) industry, it is necessary to use data mining (DM) technology to fully understand the needs of customers and make related product recommendations. Therefore, from the background of BD, this article uses DM methods to study user preferences that can increase the sales and benefits of EC companies. The main methods used in this paper are data collection method and experimental method. The application of DM in user preferences and the personalized recommendation system of EC enterprises are studied. The experimental results show that the design accuracy and coverage of this system can reach data above 0.9 under a certain threshold, but there is still an unstable phenomenon, and it is necessary to choose a good DM method for further improvement.

Keywords: Big data · E-commerce companies · User preferences · Data mining

1 Introduction

In the context of BD, the mining of user preference data is becoming more and more important. User preference refers to the user's choice and decision after subjective judgment on things or phenomena. Mining the needs and pursuits of users from massive data can effectively improve the sales performance of EC companies. Therefore, research on user preference DM is a hot topic. This can not only promote the development of EC enterprises, but also promote the best use of resources and achieve the effect of saving.

There are many theories on the application of DM in EC companies in the context of BD. For example, Han Ying said that in the new era, the rapid development of the EC industry can be said to provide EC companies with good opportunities and challenges [1]. Geng Yang explores from the perspective of EC platforms, EC users and operators, and analyzes online and offline EC [2]. Song Hao said that in the development of EC, its commercial value has gradually become important [3]. Therefore, it is necessary to study the DM of user preferences, which is beneficial to the development of business and economy.

This article first studies the theories and technologies related to BD. Secondly, it describes the function and process of DM. Then the user preferences were modeled and explored, and finally a personalized EC recommendation system was designed, and a test experiment was carried out to obtain the accuracy and average coverage.

2 Data Mining Application of E-Commerce Enterprise User Preference in the Context of Big Data

2.1 Big Data Theory and Technology

(1) BD theory

According to more and more research findings, corporate managers' demand for BD is gradually reflected in the control of competitive advantages, judgment of future trends, personalized experience of products and services, and the co-construction of corporate ecological environment. It can be seen that BD has a certain influence on the future development of enterprises. In the face of massive data, data cleaning and data sorting have become more complicated and diversified, which has given birth to the application and development of BD technology. In the era of BD, finding complete data makes real-time data extremely important. This is not only reflected in the data collection stage of the BD processing process. With the support of cloud computing, stream processing and memory analysis, a number of new algorithms have realized the real-time analysis and processing of massive data. In addition to processing massive amounts of data, companies can also optimize analysis results through incremental data. With the use of BD technology, the cost of data analysis has been reduced, and data processing resources have become more abundant. The reports given by data analysis have timeliness, continuity and virtual visibility. Enterprises give full play to the advantages of low cost of BD analysis, thereby obtaining more business model innovation opportunities and increasing the commercial value of enterprise products and services. BD has realized the efficient conversion from data to value [4, 5].

(2) BD technology

Companies can use BD as a massive information asset, and we can also artificially think that BD is data with huge scale and complex format. From a technical point of view, BD can generally use a computing framework with a hybrid architecture of Hadoop and MPP distributed databases. BD technology has the characteristics of intelligence, self-learning, visualization, virtualization, and dynamic distribution. BD can integrate data, so that under specific circumstances, data modeling can be used to simulate and predict the purchase habits of customers under different variables. BD technology can analyze the reasons for customer loss, analyze customer behavior, analyze cost composition, conduct financial analysis and risk prediction, and then based on the analysis results, companies can provide professional and personalized products and services, using their own data accumulation and external mass data, Carrying out targeted precision marketing, so as to meet the individual needs of customers at this stage with the relatively low cost of data analysis in the era of relatively small data, and further realize the economic value and social value of the enterprise. BD technology has become faster and more accurate in processing massive data, and the cost of data analysis is lower. BD visualization technology makes the operation and decision-making process of enterprises more streamlined, accurate and visual. The introduction of BD technology can make companies more efficient in resource allocation. Dynamic real-time data analysis can

enable companies to balance the ratio of production and sales, reducing or even avoiding business risks and product backlogs. On the whole, the impact of BD technology on the business model of enterprises is all-round, involving all aspects of the business model of enterprises [6, 7].

2.2 Data Mining

(1) DM function

DM is used to discover implicit and meaningful knowledge from the database. It can discover broad knowledge through concept description, related knowledge through association analysis, classification knowledge through classification and grouping methods, prediction knowledge through prediction methods, and gap knowledge through gap detection [8, 9].

(2) DM process

- 1) Ask questions and clarify assumptions.
- 2) Data collection. The first is when the data generation process is under expert control. In the second case, when experts cannot influence the process of data generation, this method is called the “observation method” [10, 11].
- 3) Data preprocessing. In the observation environment, data is usually collected from existing databases, data warehouses, and data markets. Data preprocessing usually includes at least two common tasks.
- 4) Model evaluation. Choosing and implementing the appropriate DM technology is the main task at this time.
- 5) Explain the model and draw conclusions. In most cases, DM models should help decision making.

(3) Application of association rules in user preferences

Association rules $P \Rightarrow Q$ can be described in the form of expressions, where P and Q are both non-empty itemsets. Two indicators are used to measure whether the rules are interesting. The support of the rule $P \Rightarrow Q$ is defined as the percentage of transactions $P \Rightarrow Q$ included in the database S. It can be expressed as:

$$Support(P \Rightarrow Q) = g(P \cup Q) \quad (1)$$

The confidence of the rule $P \Rightarrow Q$ is defined as the percentage of transactions that contain both P and Q in the transaction database S:

$$Confidence(P \Rightarrow Q) = G(Q|P) = \frac{support_count(P \cup Q)}{support_count(P)} \quad (2)$$

Determine whether the items are related according to the confidence of the association rules. That is to judge whether the user's preferences are directly related from the record of the user's online purchase of goods.

2.3 User Preference Modeling

In the context of the rapid development of big data, user preference data mining, as an important e-commerce product development direction, has attracted more and more attention. There are still many problems in e-commerce websites due to the fact that the e-commerce market in our country is not yet fully mature and the analysis of user behavior is not perfect. So how do we get user preferences?

(1) How to obtain

Explicit collection refers to a process that requires the use of user input to build a model to capture user preferences. Behavior tracking requires users to log in for real-time behavior interaction. Log crawling evaluates the pages that users are most interested in by analyzing the number of clicks on the website. Although this method cannot get complete information, it can still get a lot of valuable things from it [12].

(2) Influencing factors

There are many factors that affect user preferences, such as the environment, social networks, and perceived risks. The most important of these is cost. 1) Economic capacity. A person buying goods will consider his own economic strength and debt capacity. So before shopping, you must consider the issue of funds before proceeding to the next step. It is also necessary to understand the benefits of the product to consumers. This is one of the prerequisites for online shopping. 2) Age, gender, etc. Different ages and genders have different things you want to buy. Young people are more seeking brands, and older people are more seeking benefits and health. Boys like to buy shoes, girls like to buy cosmetics, clothes and so on.

(3) User modeling

1) User manual modeling

Users actively participate in modeling, enter information that expresses the characteristics of the content or fill in a personal preference form, and view topics of interest to them. The user participated in the entire modeling process. This method is used when there is more demand for personalized services.

2) User example modeling

Compared with manual modeling, this method reduces user involvement and the degree of need. After browsing the page, indicate whether the user is interested in the page and how much he likes it. The system then collects the comment pages and extracts keywords to create a user preference model.

3) Automatic user modeling

Many websites are currently using this modeling method. According to the implicit exploration method, the user's online behavior and online content are tracked, and the user's preferences are extracted.

2.4 E-Commerce Personalized Recommendation System

(1) Classification

Due to different personalized requirements, the types of EC recommendation systems are also different. According to different levels of personalized EC recommendation systems, they can be classified as follows:

- 1) Non-personalized recommendation system: The recommended content is the same for each user.
- 2) Semi-personalized recommendation system: Recommendations are generated based on different behavior records of each user. This type of recommendation system generates a recommendation set based on the user's purchase record information or the user's browsing record information. The technology usually used is the association rule, and the content recommended to each user is also different.
- 3) Fully personalized recommendation system: The system will automatically save different historical information of each user, such as registration, rating, browsing information, etc. The recommendation system will combine the historical information between the current user and other users for analysis and comparison, and finally produce a completely personalized recommendation result. This type of recommendation system is the most complete and has the highest degree of personalization.

(2) Effect

EC recommendation system can provide customers with product information intelligently. It is a new form of service that can simulate the process of sales staff helping customers complete the purchase. Its functions include: promoting website viewers to become buyers, enhancing user loyalty of EC sites, enhancing cross-selling capabilities between various sites, and optimizing EC sites.

(3) Recommended system design

The mining part is performed offline and is mainly divided into two stages. The first stage is data preprocessing. The raw data passes through the data preprocessing system. After data cleaning, user identification, session recognition, path supplementation and other steps, the original large amount of rough data is sorted into clean, accurate, and helpful data for mining. Then store it in the database. The second stage is pattern mining. The classification system classifies the data in the database, and then stores the mined patterns in the pattern library.

The recommendation part is carried out online and is mainly divided into three parts: data collection, pattern matching, and pattern recommendation. Data collection is to collect and store the interactive information between users and EC sites in the database. Pattern matching is to match the current user's behavior model with the model in the pattern library according to the user's browsing behavior information, so as to find the best pattern. Finally, the recommendation system makes intelligent recommendations to the user according to the matched pattern. The specific process is shown in Fig. 1:

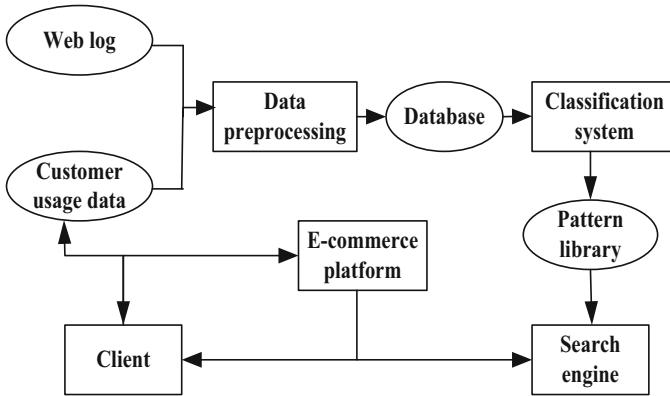


Fig. 1. Recommendation structure of E-commerce system based on data mining

3 E-Commerce Personalized Recommendation System Test

3.1 Test Data Set

The data set is the data set used to train and test the new system. As actual system data, it is important to verify and verify data records before actual system use. The type of data set depends on the type of application and the selected domain. The data selected for testing in this article is information from users who purchased books online. Therefore, the data comes from the book cross. The demonstration of the data set provides graphical information, and the data part of user privacy is anonymous.

3.2 Recommendation System Evaluation Indicators

Accuracy, the accuracy of scoring prediction is calculated by the root mean square error. TopN is a personalized recommendation list provided to users when the website provides recommendation services. Coverage describes the ability of the recommender system to extract items.

3.3 Test Process

Click the settings menu, set User Similarity Threshold to 0.8, Item Similarity Threshold 0.8, Predicted Scoring Threshold 1, Scoring Matrix Sparsity 0.79, nearest neighbor number 10, item category number 10, user category 10, recommended N Top 3, and save the settings. Then click on the recommended test, select the predictive analysis method, enter the user ID and project number, and click on the project's estimated score. In order to verify the accuracy of the association rule algorithm that combines item pooling and user pooling, various thresholds are set to test the accuracy and average coverage of the test recommendation system.

4 Test Analysis

4.1 Accuracy and Average Coverage Data Analysis

According to the test results of the system, this paper obtains the accuracy and coverage under different thresholds. This article uses the first 8 sets of data for analysis, and the specific data is shown in Table 1:

Table 1. Accuracy and average coverage data

	Accuracy rate	Average coverage
0.1	0.91	0.95
0.2	0.87	0.88
0.3	0.76	0.81
0.4	0.72	0.74
0.5	0.68	0.67
0.6	0.61	0.64
0.7	0.54	0.57
0.8	0.49	0.5

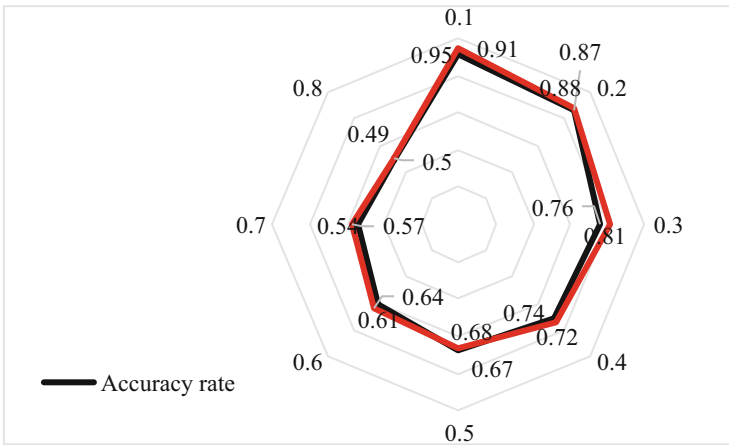


Fig. 2. Accuracy and average coverage data

As shown in Fig. 2, we can see that the highest accuracy is 0.91 and the lowest is 0.49, and it decreases as the threshold continues to rise. The average coverage rate reached 0.95 at the highest and 0.5 at the lowest, and it continued to decrease as the threshold changed. Accuracy is also related to changes in coverage, and the two complement each other.

5 Conclusion

When DM studies user preferences, it mainly analyzes and processes massive and regular data based on a large amount of incomplete and unstructured information. In the context of BD, EC companies conduct effective demand surveys on massive commodities and consumer consumption intentions, discover potential problems that may exist, and then determine marketing strategies based on this information. Through these means, the healthy development of EC companies can be promoted. According to the system designed in this paper and the test results, it is found that the accuracy and coverage of the system designed in this paper have a strong correlation with the threshold, so it is necessary to improve the function of the system to ensure the accuracy and coverage of the system.

References

1. Han, Y.: Analysis of online and offline e-commerce user data mining process based on big data technology. *Dig. World* (10), 134–134 (2018)
2. Geng, Y.: Research on data mining of online and offline e-commerce users based on big data (2014-18), 119 (2021)
3. Song, H., Jiang, F.: Data mining of online and offline e-commerce users based on big data. *Electron. Technol. Softw. Eng.* (16), 160–160 (2017)
4. Popchev, I., Ketipov, R., Angelova, V.: Risk averseness and emotional stability in e-commerce. *Cybern. Inform. Technol.* **21**(3), 73–84 (2021)
5. Linda, S., Minz, S., Bharadwaj, K.K.: Fuzzy-genetic approach to context-aware recommender systems based on the hybridization of collaborative filtering and reclusive method techniques. *AI Commun.* **32**(2), 1–17 (2019)
6. Lytvyn, V., Vysotska, V., Demchuk, A., et al.: Design of the architecture of an intelligent system for distributing commercial content in the internet space based on SEO-technologies, neural networks, and machine learning. *East. Eur. J. Enterp. Technol.* **2**(2(98)), 15–34 (2019)
7. Anaya-Sánchez, R., Castro, M.: Millennial consumer preferences in social commerce web design. *Revista Brasileira de Gestao de Negocios* **22**(1), 123–139 (2020)
8. Yadav, V., Shukla, R., Tripathi, A., et al.: A new approach for movie recommender system using K-means clustering and PCA. *J. Sci. Ind. Res.* **80**(2), 159–165 (2021)
9. Palomares, I., Porcel, C., Pizzato, L., et al.: Reciprocal recommender systems: analysis of state-of-art literature, challenges and opportunities towards social recommendation. *Inf. Fusion* **69**(16), 103–127 (2021)
10. Bertani, R.M., Bianchi, R., Costa, A.R.: Combining novelty and popularity on personalised recommendations via user profile learning. *Expert Syst. Appl.* **146**, 113149.1–113149.8 (2020)
11. Prastawa, H., Ciptomulyono, U., Laksono-Singgih, M., et al.: The effect of cognitive and affective aspects on usability. *Theor. Issues Ergon. Sci.* **20**(1), 1–27 (2019)
12. Kremez, Z., Frazer, L., Quach, S., et al.: Collaboration, communication, support, and relationships in the context of e-commerce within the franchising sector. *J. Strateg. Mark.* **101**, 1–23 (2020)



Analysis of English Teaching Data on Cloud Platform Based on Data Mining Technology

Zhuolin Zhou(✉)

Guangxi Agricultural Vocational and Technical College, Nanning 530022, Guangxi, China

Abstract. In vocational education, English education plays an increasingly prominent role and significance. Analyzing the data generated in English education through data mining technology can help teachers carry out targeted teaching better and help students find problems in learning. This paper focuses on the impact of data mining on College English teaching. Combined with the data model of the cloud platform, four types of data are extracted from the platform, namely the initial scores of the students, the test scores of the students after the experiment, the control group and the experimental group before and after the test, the degree of preference for English courses, learning interest, independent learning ability and cooperative learning. The survey data of ability, etc., and the survey data of the experimental group's adaptability to the teaching model of "learning by learning" supported by educational big data. Classroom teaching observation provides a guarantee for the reliability of the experiment. During the observation, the subjects were not informed in advance. Therefore, the subjects behaved naturally, reflecting the real phenomenon of ordinary English listening classes. Under the assumption that the variances are equal, the P value of the mean T test is 0.117 far greater than 0.05, indicating that there is no significant difference in the performance of the two classes. The results show that improving the college public English teaching syllabus and formulating college public English education policies that conform to the trend of the times have very important practical significance.

Keywords: Data mining · Data analysis · College English · Teaching data

1 Introduction

China has entered an era of rapid development, and its economic and cultural development is at the advanced level in the world. It will also bring a variety of contradictions and problems, and complex. As college students in this new era, their world outlook, outlook on life and values have not yet been formed. If they do not have the correct concept of right and wrong and the ability to distinguish, they can not distinguish a variety of mutual impact values, and can not make the right choice and choice.

This paper investigates the current situation of professional English Teaching in Colleges and universities, and studies the teaching plan of professional English reform according to the theoretical guidance of pedagogy, so as to make it meet the needs of

agricultural development [1]. It also puts forward that in the process of professional English teaching reform in higher vocational colleges, new methods should be used to improve the teaching level of professional English and make students have good quality, So as to cultivate compound talents [2, 3]. There is a close relationship between language and culture, and the differences of knowledge in different cultural backgrounds will lead to students' difficulties in oral output [4, 5]. Only by understanding the cultural background of these countries can we have a profound understanding of their life style, religious belief, way of thinking, values and aesthetic taste [6, 7], which will also help to improve their language ability [8]. Combined with the analysis of relevant scales, this paper discusses the value of this model in stimulating students' learning motivation and improving academic performance, and demonstrates the feasibility and necessity of this model [9, 10].

As an important part of China's education system, vocational education itself has strong practical requirements, but in the actual teaching work, it plays a relatively small role in improving students' practical ability. Through the current English teaching reform in higher vocational colleges, we can solve the problems of passive learning and passive learning of higher vocational students, and let students master the learning mode of integrating professional skills and English knowledge. To a great extent, it can enhance the enthusiasm of students to participate in English practice teaching, improve the participation of vocational students in English learning, and effectively improve their learning efficiency.

2 English Teaching in Colleges and Universities

2.1 Data Mining Technology

In the Internet, each user can create and send their own information through the human-computer interface, forming a flow of information, which can be transferred, processed, analyzed and stored in the Internet through the invisible flow way. This is a kind of information exchange with individual users as the information source; other users who can establish a network connection relationship with the information source in the Internet can receive this information through the information presentation of the information flow interface. The traditional teaching mode mainly focuses on Teachers' explanation and students' acceptance. Teachers play a leading role while students learn passively. Education big data makes it possible for teachers to obtain all kinds of process data information of each student. Teachers can understand the characteristics of each student and have insight into the needs of all students under the support of platform data, so as to effectively guide students' learning, diagnose and solve students' problems, and truly realize personalized learning.

The motion equation of the general linear model multi-agent system is:

$$\dot{x}_i(t) = Xx_i(t) + Yu_i(t), i = 1, 2, \dots, n \quad (1)$$

The specific form of control input is:

$$u_i(t) = - \sum_{j \in N_i(t)} \nabla_{q_i} \psi_\alpha(\|q_i - q_j\|_\sigma) - \sum_{j \in N_i(t)} a_{ij}(t)(p_i - p_j) + c_1(q_r - q_i) + c_2(p_r - p_i) \tag{2}$$

Among them, $c_1, c_2 > 0$, q_i and p_i are the position vector and velocity vector of agent i , respectively.

In order to generate repulsive force between the two agents, a new artificial potential function is redefined, and its expression is:

$$\psi_\beta(z) = \int_{d_\beta}^z \phi_\beta(s) ds \tag{3}$$

Among them, $\phi_\beta(z)$ is the force function.

2.2 Industry English

In the whole process of education, we need to integrate the relevant professional vocabulary and professional knowledge with the major. It is also a feasible way for the public English teaching objectives to meet the needs of social talents. Through English practice teaching, it can effectively provide a broader development space for higher vocational students. Teachers’ professional development is not only teachers’ personal behavior, because it is not the ability of individual teachers to achieve, so we should start from the school level, realize mutual learning and unified management, at the same time, all kinds of learning and training need financial support, which also requires the school leaders to adjust the management mode and give certain financial support.

3 Industry English Teaching Experiment

3.1 Data Collection

Supported by education big data, the teaching mode of “teaching by learning” has been practiced in junior high school English class for one semester, and about 100t of data has been stored on the cloud platform. Combined with the data model of cloud platform, four types of data are extracted from the platform, which are the students’ initial scores, students’ test scores after the experiment, the survey data of the control group and the experimental group on their liking for English courses, learning interest, autonomous learning ability and cooperative learning ability, and the survey data of the experimental group on the adaptability of the teaching mode of “teaching by learning” supported by big data.

3.2 Statistical Analysis

Classroom teaching observation provides a guarantee for the credibility of the experiment. The observation did not inform the subjects in advance, so the subjects behaved naturally, reflecting the real phenomenon of English listening class.

4 Discussion

4.1 Teaching Effect

The difference test of the final scores between the control class and the experimental class is shown in Table 1. Under the assumption of equal variance, the p value of meant test is 0.117, which is far greater than 0.05, indicating that there is no significant difference between the two classes. But the average score of the students in the control class is 75.31 and the average score of the students in the experimental class is 81.28. Enterprises and training classes should participate in practice, integrate professional knowledge in the process of language learning, and attach importance to practice in the process of classroom teaching, so as to help students and solve their practical problems in English language application.

Table 1. Difference test

Class	N	Mean value	Standard deviation	Standard error of mean
Pre test of experimental class	51	71.780	15.559	2.196
Post test of experimental class	51	81.280	12.659	1.796

4.2 Analysis of Teaching Mode

The students' interest in learning English is shown in Fig. 1. From the figure, we find that some vocational college students are willing to learn English subjectively, and realize the importance of English learning, which has practical significance in academic and employment. But there are still 18% of the students do not want to learn English, completely lost interest. There are significant factors in English achievement and auxiliary tool of film and television materials in English learning, film and television resources in English teaching and the improvement of English teaching environment, and English learning methods. There are no significant differences between gender and instrumental factors of film and television materials in English learning, between film and television resources in English teaching and the improvement of English teaching environment, and between ways and methods of English learning. There is no significant difference in the auxiliary tool factor of film and television materials and the method factor of English learning between major and English learning, but there is significant difference in the promotion factor of English teaching equipment and environment. The scientization of college foreign language education policy is reflected in the fairness of foreign language education policy-making process, the rationalization of language planning and the conformity of curriculum with the law and performance of language development. In different links, modern science and technology should be used as far as possible to ensure that the operation and function of each link can be brought into full play. This is an education system based on the socialist planned economic system and social management

mode. In the early stage of large-scale economic construction, this strict government plan can better ensure that education is in line with the development of economy and society, so as to better realize the goal of education public interest.

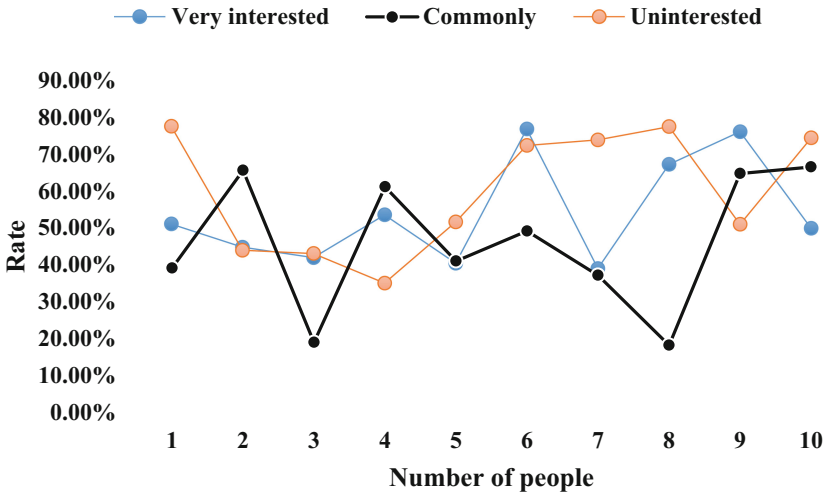


Fig. 1. Students' interest in learning English

The time for students to learn English every week is shown in Fig. 2. According to the data, 92.3% of the students study English less than one hour a day, of which 70.3% study English less than five hours a week; only 7.7% study English more than nine hours a week. These data show that students generally do not pay attention to English learning, and only less than one tenth of students are willing to spend time and energy on English learning. Many English teachers actively learn the theoretical knowledge of English listening strategies. In English listening training, many teachers begin to enter the classroom and guide students' listening practice. And actively explore effective measures to improve students' English listening level. In addition, a campus English corner can be set up to provide students with extra-curricular activities related to English, such as watching original movies, watching English programs and listening to English songs, so that students can have more opportunities to contact English and build a platform for them to learn and use English. The anxiety of college students in the process of English learning is an important factor affecting the development of their language ability and the improvement of their English level. Therefore, teachers should not attack students, but be full of love, fully respect students' personality, try every means to stimulate students' learning enthusiasm in teaching, and strive to cultivate their self-confidence, so that students can find and play their potential.

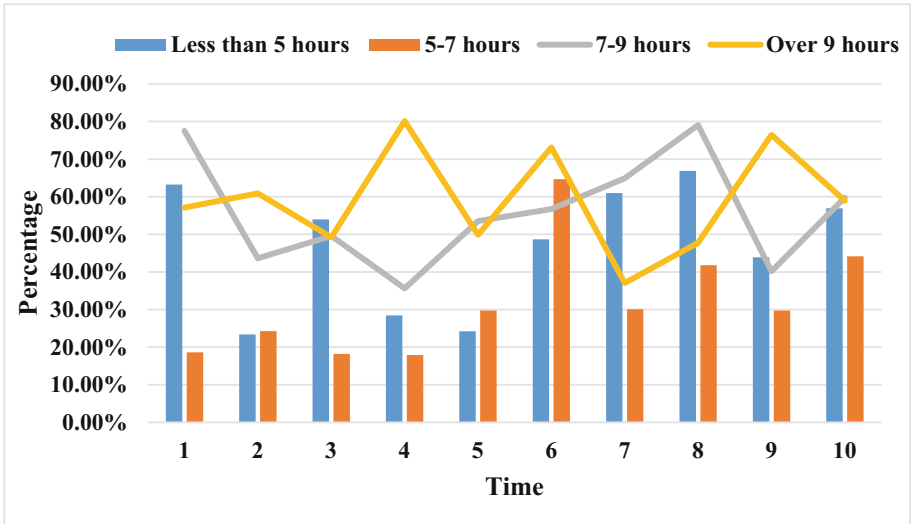


Fig. 2. Time of students learning English every week

5 Conclusions

The era of big data has brought huge opportunities and challenges to social media. Guide and promote the subject to establish reasonable value orientation, deconstruct and eliminate unreasonable value orientation. Improving the quality of English teaching and adapting to the needs of the vocational education system for talents in the new era is the fundamental purpose of the evaluation of higher vocational English teaching quality.

Schools should increase their investment in industry English teaching, update equipment, introduce modern information technology, and combine online and offline teaching, so that learning is not confined to classrooms and classrooms, and should break the time and geographical constraints to create more for students learning opportunities.

For most advanced level learners who can use English proficiently, they should choose some video materials that are highly applicable and interested in themselves, which can help them understand English culture and cultivate their English thinking style.

Acknowledgments. Guangxi Vocational Education Teaching Reform Research Project in 2019 “Practical Research on the flipping classroom Teaching Model of English courses in higher Vocational Education based on Micro course Co-construction between School and Enterprise-taking Hotel English Curriculum as an example “(item number: GXGZJG2019B085).

References

1. Moser, K., Zhu, D., Nguyen, H., et al.: Teaching English language learners: a mainstream response to rural teacher preparation. *Int. J. Teach. Educ. Prof. Dev.* **1**(1), 58–75 (2018)

2. Tenridinanti, T.B.: The Impact of a brief teaching English towards the drivers online transportation in dealing with ASIAN GAMES 2018. *Int. J. Psychosoc. Rehabil.* **24**(1), 1944–1949 (2020)
3. Alhasov, Y., Verbytska, A., Kolenichenko, T.: Teaching English to adult learners within extracurricular activities at university: barriers and motivation factors. *Adv. Educ.* **7**(15), 12–19 (2020)
4. Gryzhenko, A.: General aspects of teaching English at higher educational establishments in modern conditions. *Humanit. Sci. Curr. Issues* **1**(32), 177–182 (2020)
5. Ali, R., Abd, M., Algane, A.A.: Teaching English literacy in the time of COVID-19 pandemic in higher education: a case study in Saudi Qassim University. *Multicult. Educ.* **6**(5), 204–215 (2020)
6. Ali, N.: The effect of using technology-enhanced English course (TEEC) in teaching English as a foreign language at preparatory schools. *Int. J. Psychosoc. Rehabil.* **24**(5), 2658–2666 (2020)
7. Aini, R.M., Muslem, A., Usman, B.: Lecturers' strategies used in teaching English reading comprehension: a case at a private college in aceh. *Humanit. Soc. Sci. Rev.* **8**(3), 20–30 (2020)
8. Rosalinah, Y., Khilda, N.: Teaching English speaking through information gap technique (a classroom action research at the 8th grade students of Madrasah Tsanawiyah Al-Husna Depok). *Wanastra Jurnal Bahasa dan Sastra* **11**(2), 159–166 (2019)
9. Maldaner, L., Barbosa, M.N., Barbosa, S.: Teaching English as a foreign language in a Brazilian public school in balsas (MA) through collaborative activities. *EntreLetras* **10**(1), 136–148 (2019)
10. Abdulloh, A.: The importance of submitting mini plan before teaching English as foreign language at intensive English course Bekasi branches. *J. Engl. Lang. Lit. (JELL)* **2**(2), 107–116 (2018)



Realization of the Character Analysis Algorithm of Science Fiction Characters Based on Text Information

Ying He^(✉)

School of Translation Studies, Xi'an Fanyi University, Xi'an 710105, Shaanxi, China

Abstract. Science fiction is the most successful art form that combines science and technology with literature. It focuses on the profound impact of technological development on natural ecology and human civilization. After a long course of more than a century, Chinese science fiction has gradually evolved from naive to mature. The expansion of writing themes, the improvement of writing methods, and the continuous expansion of creators and readers have greatly improved the quality and scope of Chinese science fiction. At the same time, the development of science and technology and the introduction of western ideas such as philosophy and ecology have had a profound impact on the development and development of Chinese science fiction. This article aims to study the realization of the character analysis algorithm of science fiction characters based on text information, and to understand the dialectical understanding of the influence of Chinese science fiction on science and technology in different eras. Through a large number of literature readings and experiments, this paper has thoroughly studied the realization of the algorithm of character analysis of science fiction characters by text information, summarized the influence of science fiction characters' personalities, and used the text information mining algorithm to analyze the characters of science fiction novels, and designed science fiction. Flowchart of text mining for character analysis of novel characters. At the same time, using Chinese traditional culture, Western philosophy and other related theories to study the reasons for changes in scientific and technological concepts and make evaluations, it has certain reference value and practical significance for the study of Chinese science fiction and human social practice.

Keywords: Text information · Science fiction · Character analysis · Algorithm research

1 Introduction

Because science fiction is the literature most closely related to science and technology, and the development and progress of science and technology is the inexhaustible source of science fiction, so science fiction writers are more sensitive to the pros and cons of science and technology [1, 2]. In the past, a large number of new generations of sci-fi writers have emerged in the Chinese sci-fi world, bringing a large number of

outstanding scientific and technological works and art critics to the Chinese sci-fi world. They no longer use science as a tool to “populate scientific knowledge” and no longer use technology. Treated as an artifact. Instead, science fiction writers adjusted their “science and technology vision” to explore science fiction with a larger and more ambitious vision. Science fiction writers think about science and technology, treat science and technology dialectically, reflect on deeper character issues and the living environment of the science and technology era, and embody a strong sense of humanism [3, 4].

In the field of science fiction development, many experts and scholars have conducted a lot of research and achieved good results. For example, V Friedrich proposed that Western science fiction should also be historical and temporal. The author carefully shapes place rather than positioning in time and space, imagines the future and reshapes history in a relative historical concept. However, whether it is Both remake and destruction occur in a time and space relatively separated from the real world, which is an extension of utopian novels in the world of science fiction [5]. Persson AB believes that the difference between the new generation of Chinese sci-fi writers is that the new generation of sci-fi writers is a review of real history. The past described in the novel is not a virtual world in the back sky, but the fragments and characters of Chinese history that readers are familiar with [6].

The development of Chinese science fiction for more than a century has brought us many works, which have become the mainstream trend. Due to the technological background of social development and the unique “seeking truth from facts” nature of science fiction, the creation and development of Chinese science fiction is a scientific and technological perspective with the reality of the times and context [7, 8]. Science fiction is not only related to the construction of Chinese literary theory, it has guiding significance for our social practice, but also has a certain practical significance. This article proposes a comprehensive research method that more systematically presents the specific manifestations of scientific and technological concepts in Chinese science fiction. Combining the context of the times, it examines the different attitudes to science and technology in different development periods of Chinese science fiction, and deeply reflects and criticizes the negative impact of the development of science and technology on the natural environment, social order and character. Compared with the existing research results, this article explores the roots of the formation of scientific and technological concepts from Chinese traditional cultural thoughts, Western philosophy and ecology, and puts forward new requirements for the further development of Chinese science fiction. It is hoped that the further development of Chinese science fiction will reflect A unique and highly humanistic spirit and a focus on reality.

2 Science Fiction and Text Information

2.1 The Influence of the Character of Science Fiction

- (1) For people who were born in a natural environment and grew up in a social environment, changes in nature and society have a profound impact on everyone’s personality. From the awakening of people’s “self-awareness”, they gradually become the masters of all ecology. As science and technology become a key tool to transform the natural society, humans themselves as subjects have undergone major changes.

This change is reflected in the fact that technology has not only improved the material living conditions of the rich, but more importantly, the human spiritual world has undergone earth-shaking changes. Especially with the emergence of private property, people gradually developed “selfishness” and “greed”, exposing some ugly aspects of characters. As the industrial revolution, wars, religious reforms, and cultural ideas are rapidly pulling the gears of society, we must admit that the personalities of the characters are also changing. In this beautiful residence created by technology for us, the material level has indeed improved, but the spiritual level has brought many problems [9, 10].

- (2) Technology has brought the convenience of Internet communication, but it has made the characters more indifferent. When it comes to the relationship between technology and people, character characters seem to be particularly important. The new generation of science fiction writers believe that modern society is changing very fast, and technology plays a very important role in this process. Different character characters represent different people’s psychology. The technological changes in people’s minds are very limited in traditional literary writing, so a new literary form is needed to describe the current state [11, 12].
- (3) Science and technology have brought rationality to human development from a critical perspective. At the same time, from the perspective of reality and humanity, pay attention to the impact of technology on human nature. The development of science and technology has given birth to the arrival of a networked information society, and the application of networked information technology has greatly affected our way of life and work. People bathed in the “information ocean” can use their minds more flexibly and quickly hold wealth and power firmly in their hands.
- (4) With the advent of computer networks, the world has not only become smaller, but also more colorful. Network information technology has played an indispensable role in the development of society, politics, economy and culture. For individuals, it broadens people’s horizons, broadens the platform for people to communicate with each other, and creates a new way of virtual existence. Through the Internet, people can cross the boundaries of time and space and communicate with others anytime, anywhere. In the era of information explosion, the Internet has become an indispensable device for netizens. However, if we are immersed in the beauty of the Internet, we cannot ignore its impact on the other side of our lives.

2.2 Gini Index Algorithm Based on Text Mining

In order to study the character of science fiction characters in depth, this paper uses the Gini index algorithm of text mining in text information to analyze the character of science fiction characters. A basic method for selecting the most typical attributes of a document category from a large number of science fiction character data text sets is to evaluate the selected text of the science fiction character character. As far as the effect of character classification in science fiction is concerned, one of the most commonly used methods to quantify the effects of various characteristics of science fiction characters is the measurement of the Gini index. Suppose that word A has a label of type F , and the corresponding weight is $T_1(A), \dots, T_F(A)$, that is, the higher probability that the document

belongs to the n th category is $T_n(A)$. Therefore, the Gini index of word A is represented by $S(A)$ as follows:

$$S(A) = \sum_{n=1}^F T_n(A)^2 \quad (1)$$

$$T'_n(A) = \frac{T_n(A)/T_n}{\sum_{m=1}^F T_m(A)/T_m} \quad (2)$$

Among them, $\sum_{n=1}^F T_n(A)^2 = 1$, the value range of the Gini index $S(A)$ for character analysis of science fiction characters is in the range of $(1/F, 1)$. If the value of $S(A)$ of the character data of science fiction characters is larger, it means that the resolution of the word F is higher. The basic idea of the Gini index algorithm for analyzing the character text mining of science fiction characters is to evaluate the impact of each text element of the character analysis of science fiction characters on the character classification performance of the character analysis of science fiction, the higher the purity of the text classification of the character analysis of science fiction characters, Which means that the better the attributes of the character data of science fiction characters are selected for analysis.

3 Experimental Simulation

3.1 Experimental Background

In order to study the character of science fiction characters in depth, this paper uses the text mining technology of text information to analyze the character of science fiction characters, and designs the flow chart of the text mining system of science fiction characters. The text mining of science fiction character character analysis refers to extracting meaningful and valuable science fiction character character information from a database of many unstructured science fiction character character text resources to identify patterns such as classification and grouping. Generally speaking, it is to make computers recognize natural language like human beings. The general process of text mining technology is shown in Fig. 1:

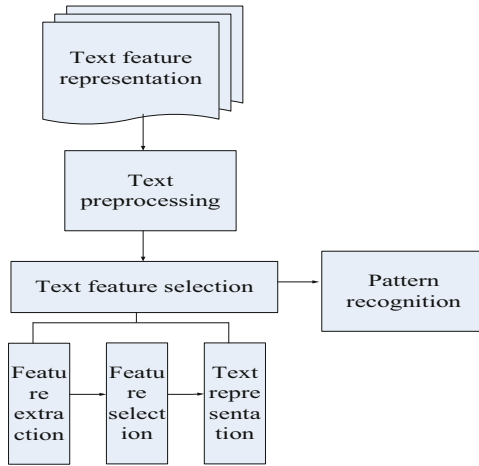


Fig. 1. Basic process of text mining

Figure 1 is a flow chart of the character text mining system for science fiction characters, including text preprocessing, feature separation, feature selection, text rendering, pattern recognition and so on. The text preprocessing of the character data of science fiction characters is mainly to filter out irregular or useless information in the character data of science fiction characters, so as to provide a prerequisite guarantee for the identification step of the character analysis mode of science fiction characters. After preprocessing the character data of science fiction characters, it is generally the text characteristics of the character data of science fiction characters. Textual demonstration of character data of science fiction characters. Extracting data from science fiction characters is to transform unstructured information into structured vectors. Before the presentation of the character data of science fiction characters, science fiction needs to go through the function selection. The function selection of the character data of science fiction refers to the removal of some useless data to reduce the dimensional characteristics of the character data of science fiction characters.

4 Experimental System Data Test

4.1 System Adaptation Test Parameter Analysis

The following are three experimental groups (Group 1, Group 2, Group 3) tested the data of the adaptability of the science fiction character text mining system A1 ~ A3. The experimental data is shown in Table 1.

Table 1 is the simulation parameter table of the adaptation test of the science fiction character text mining system. From the table, we can see the science fiction character text mining system A1–A3 tested by the three experimental groups (Group 1, Group 2, Group 3) The degree of fitness. The adaptation test of the science fiction character text mining system is to test the adaptation degree of the science fiction character data and the system. The larger the test data, the higher the adaptation degree of the science fiction

Table 1. System adaptation test parameter setting

Fit test	Group 1	Group 2	Group 3
A1	0.66	0.85	0.75
A2	0.98	0.61	0.63
A3	0.74	0.79	0.84

character data to the system. The data of the character text mining system for science fiction characters are all above 0.6, indicating that the adaptation test of the character text mining system for science fiction characters has a high degree of adaptation to the system.

4.2 Analysis of System Safety Test Parameters

As shown in Fig. 2, the following is the safety test data of the science fiction character text mining system F1–F3 tested by three experimental groups (Group 1, Group 2, Group 3).

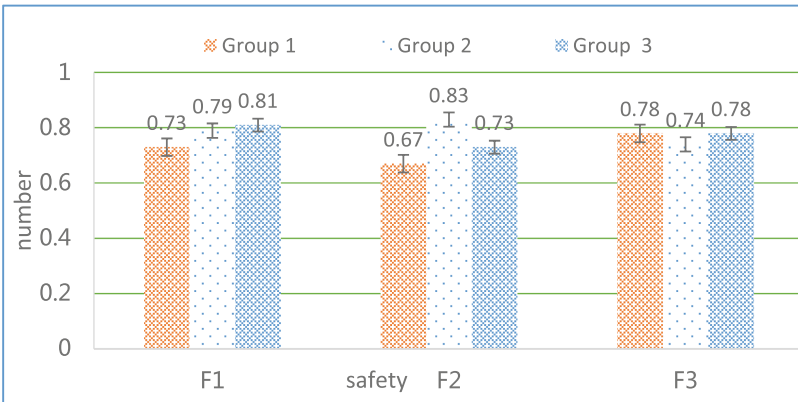


Fig. 2. System safety test parameter analysis

Figure 2 is a safety test parameter diagram of the F1–F3 character text mining system for science fiction characters. From the figure, we can see the security situation of the character text mining system F1–F3 of science fiction characters. Among them, in the security of the science fiction character text mining system tested by the text mining system, the F1 system has a security of 0.73 for the first group, 0.79 for the second group, and 0.81 for the third group; with a text mining system The security of the tested science fiction character text mining system. The F2 system has a security of 0.67 for the first group, 0.83 for the second group, and 0.73 for the third group; the character of a science fiction character tested by the text mining system In the security of the text mining system, the F3 system has a security of 0.78 for the first group, 0.74 for the

second group, and 0.78 for the third group. The three sets of security data in the security of the science fiction character text mining system tested with the text mining system are higher, indicating that the system's data is safer.

5 Conclusions

Science fiction is a kind of “newborn” novel, which is closely related to real life like other literary genres; the difference is that science fiction is a science-based and evolving imaginative art, and its historical process may involve scientific knowledge. Popularize and use scientific logical reasoning. The characters in science fiction have distinctive characteristics, lively and active, and they have intelligent brains. Through a large number of literature reading and experiments, this article has done an in-depth study on the realization of the character analysis algorithm of science fiction characters based on text information, and summarized science fiction. The influence of the character's personality, the Gini index algorithm of text mining in the text information is used to analyze the character of science fiction characters, and the flow chart of the text mining system of science fiction character's character is designed. Although science fiction is not a professional book to explain scientific knowledge, it can arouse people's interest in science, promote scientific innovation, and promote the development of science; it can also inspire people to correctly understand the shortcomings of technology, use technology correctly, and view the impact of technology accurately; Even to a certain extent, science fiction can give people dreams, enlightenment and power. Therefore, it is particularly important to promote the development of science fiction in the Chinese literary world.

Acknowledgements. Project supported by the Research Foundation of Education Department of Shaanxi Province (Grant no. 16Jk2058).

References

1. Emanuelle, B., Judy, G., Nicholas, M.: How to teach computer ethics through science fiction. *Commun. ACM* **61**(8), 54–64 (2018)
2. Avila-Bront, L.G.: An experiential learning chemistry course for nonmajors taught through the lens of science fiction. *J. Chem. Educ.* **97**(10), 3588–3594 (2020)
3. Gough, N.: Specifying a curriculum for biopolitical critical literacy in science teacher education: exploring roles for science fiction. *Cult. Sci. Edu.* **12**(4), 769–794 (2017). <https://doi.org/10.1007/s11422-017-9834-0>
4. Gleaves, J.: Beyond prometheus, strawmen, and science fiction: ethicists and the moral debate over enhancements to human performance. *Kinesiol. Rev.* **6**(1), 91–98 (2017)
5. Friedrich, V.: Die Rhetorik der Technik als Motor der Utopie: ber den bedeutenden deutschen Science-fiction-Roman Auf zwei Planeten von Kurd Lawitz. *Rhetorik* **39**(1), 15–33 (2020)
6. Persson, A.B.: Home and hell: representations of female masculinity in action-driven science fiction literature. *Lambda Nordica* **25**(2), 68–90 (2020)
7. Manase, I.: Science fiction and the imagined South African and other worlds in Lauren Beukes's novels. *Imbizo* **10**(2), 1–17 (2020)

8. Castleman, B.L., Meyer, K.E.: Can text message nudges improve academic outcomes in college? Evidence from a west virginia initiative. *Rev. High. Educ.* **43**(4), 1125–1165 (2020)
9. Mccarthy, O.L., Wazwaz, O., Jado, I., et al.: An intervention delivered by text message to increase the acceptability of effective contraception among young women in Palestine: study protocol for a randomised controlled trial. *Trials* **18**(1), 454–456 (2017)
10. Burn, E., Nghiem, S., Jan, S., et al.: Cost-effectiveness of a text message programme for the prevention of recurrent cardiovascular events. *Heart (Br. Card. Soc.)* **103**(12), 893–894 (2017)
11. Rand, C.M., Vincelli, P., Goldstein, N.P.N., et al.: Effects of phone and text message reminders on completion of the human papillomavirus vaccine series. *J. Adolesc. Health* **60**(1), 113–114 (2017)
12. Kerbaj, J., Toure, Y., Aladro, A.S., et al.: Smartphone text message service to foster hand hygiene compliance in health care workers. *Am. J. Infect. Control* **45**(3), 234–239 (2017)



Application of Mass Data Retrieval Technology Based on Cloud Computing

Ruijuan Li¹ and Shunji Zhang²(✉)

¹ Qujing Medical College, Qujing 655000, Yunnan, China

² Qujing Normal University, Qujing 655000, Yunnan, China

Abstract. As an emerging computing model, cloud computing has developed rapidly in recent years due to its convenient management, low cost, dynamic growth of storage space and no need to purchase storage equipment. The purpose of this article is to study the application of massive data retrieval technology based on cloud computing. First, it introduces the field of cloud computing data storage and retrieval mechanism in three aspects, including data storage technology, data organization and management, and data retrieval technology. After that, the basic idea of solving the problem in this article is put forward, namely, the construction environment of cloud computing, and then the solution is introduced by establishing a keyword-based content retrieval index model. Finally, the effectiveness of the algorithm is verified by comparing it with multiple algorithms on different data sets. If the index is established, the average insertion speed is about 162 m/s, and the average single-node insertion speed is 27 m/s due to Observer interception, index creation, and more time for index data to be inserted into the table. That is to say, indexing makes each node the average insertion speed is 3 m/s slower.

Keywords: Cloud computing · Massive data · Retrieval technology · Cloud database

1 Introduction

The development of the Internet makes network security attacks happen frequently, and network spam also occupies a certain proportion of storage space [1]. In traditional databases, in the face of the rapid increase in data volume, people have adopted different strategies to deal with them. Cloud database is a general term for the storage services provided by cloud computing for users. It has the advantages of cloud computing and, compared with traditional relational databases, the advantage of cloud database is: from the perspective of storage space, a large table stored in a traditional database is divided into several small tables and stored in each sub-node [2, 3]. From the perspective of query time, the same query request can be executed in parallel in multiple child nodes, thereby reducing the time required for query. With the rapid development of cloud computing technology after its emergence in China, companies have seen the advantages of cloud databases and deployed cloud database architectures such as Baidu Cloud and Huawei Cloud [4, 5].

In the cloud computing environment, the retrieval of massive ship fault data is easily interfered by association rule items, and the fuzzy clustering of data retrieval is not good. In order to improve the efficiency of ship fault diagnosis, Zhang S invented a complex query support based on the cloud computing environment Retrieval technology of massive ship fault data. The vector quantization feature coding technology is used to analyze the distributed storage structure of the massive ship fault data, and the segmented adaptive regression analysis method is used to analyze the frequency spectrum feature of the massive ship fault data. The quantitative recursive analysis model is used to extract the massive ship fault data [6]. Internet person searches usually return pages related to several different names with the same name. Wei Y proposes a new web model for character data extraction without template, and uses the Dirichlet Process Mixture model to solve the name disambiguation problem. The results show that their method works best on web pages with complex structures [7]. Therefore, it is of great practical significance to assist indexing for data retrieval technology on massive data [8].

Based on the research results of real-time database systems and the core technology of cloud computing, this paper focuses on solving the problems of massive real-time data storage, retrieval, backup data distribution, consistency repair, and system dynamic expansion in a distributed environment. It is based on the existing real-time database technology on the basis, after transformation and innovation, the “distributed real-time database data storage and retrieval mechanism based on cloud technology” has been researched and realized. Finally, this paper designs experiments to verify the performance of retrieval using RowKey as the primary key, the establishment of secondary indexes, and the performance of Impala and Hive data query, and the comparison and analysis of the results confirm that the above design ideas can be achieved by the massive data retrieval technology of cloud computing And better than expected goals [9].

2 Research on Retrieval Technology of Massive Data Based on Cloud Computing

2.1 Cloud Computing Data Storage and Retrieval Mechanism

(1) Data storage technology

Data storage mainly implements functions such as distributed storage and redundant backup of real-time data sent to the server cluster. High-performance data storage divides, isolates, and abstracts distributed storage into interfaces and single points by using and distributed communication service platform [10].

The snapshot subsystem realizes the rapid reception of data sent from multiple collectors to the server cluster. After unpacking, the control group state information obtains its corresponding bit number and then forwards the data to the designated server node or performs local processing [11, 12].

(2) Data organization and management

The organization and management of data sets distributed characteristics and real-time requirements, and designs the data organization and management plan in the system. The design is mainly divided into three parts.

The data point identification design realizes the combination of global management and local management of data points, and can quickly realize data positioning.

The design of the data recording unit has designed the smallest unit of system data operation in accordance with the characteristics of real-time data.

The “versioned” data management mechanism is a mechanism suitable for distributed real-time data storage and retrieval based on the design of data points and data recording units, and assists in multiple redundant backups and data consistency restoration.

(3) Data retrieval technology

Data retrieval mainly realizes the efficient retrieval of data under the distributed architecture. It uses the distributed communication service platform to distribute retrieval tasks and isolates complex associations between nodes. The research focus is transformed into current node data retrieval, which is mainly divided into data retrieval algorithm data retrieval research. The work mainly includes two parts: data query algorithm design and cache area management.

The data retrieval algorithm has designed a data retrieval structure and query algorithm for the “versioning” of the data, so that the data retrieval has good real-time performance.

The cache area management selects the page replacement algorithm and sets the parameters according to the system query characteristics to ensure a high hit rate and improve query performance.

2.2 Cloud Computing Environment Construction

(1) Hadoop distributed cluster construction. In the cloud computing environment, the master node master is used to manage each sub-node slave, and the NameNode running on the master node manages and controls the DataNode of each sub-node. The test environment of this article is three virtual machines with ubuntu11.10 operating system built on a WindowsXP computer. Under the initial conditions, one virtual machine is used as the master (hostname is named ubuntu11), and two virtual machines are used as child nodes (the hostname is named ubuntu112 and ubuntu113 respectively).

Installing the Hadoop cluster cloud environment generally includes installing JDK, installing SSH to achieve passwordless access between master and slaves nodes, and

installing Hadoop. Considering the stability of Hadoop, this article uses version 0.20.2 and its corresponding JDK version 6u20-linux-i586.bin.

2.3 Keyword-Based Content Retrieval Index Model

There are only two data retrieval methods in HBase. One is to execute the query through the get party based on the known key information, and the other is to traverse the data table and obtain the data without knowing the key information through the scan method. If you search based on keywords in the HBase database, you can only match one by one through the scan method, which is an extremely time-consuming operation in a massive data environment. This article uses another form to realize data query for keywords. First obtain the user's input sentence content, and divide the content into keywords through the Analyzer; then query the index document according to the or relationship between the keywords. If there is no corresponding record, it will directly return to inform the user that there is no relevant information, otherwise Search for documents named with keywords to obtain the corresponding RowKey information, thereby obtaining complete data information.

In the HBase cloud database, the data is stored in the distributed file system HDFS, and the query efficiency for RowKey is very high. However, the query for non-RowKey needs to traverse the data table, and the query efficiency is extremely low. In order to improve the efficiency of keyword retrieval for the Weibo content stored in HBase, the main stages of semantic analysis, index construction, index storage, and query execution based on the index are added. Although these four stages will increase additional overhead, the query can be converted into a query that obtains tuple information based on RowKey, and the four stages only need to be established once, and the index maintenance of existing data will no longer be involved in future queries related work. In the index maintenance process, each child node maintains the data stored on it, that is, Region and Index have a one-to-one relationship.

3 Investigation and Research of Mass Data Retrieval Technology Based on Cloud Computing

3.1 Experimental Procedure

To test the data insertion performance before and after the establishment of the secondary index for HBase, the experiment uses two clients, and the amount of data used is 10G, 20G, 30G, and 40G. At this time, Memestore is set to 128M by default. The experimental steps are as follows: continuously insert log data into the first client; continuously insert data into the second client and enable the coprocessor to create indexes; record the time it takes for each amount of data to be inserted separately; draw based on the above data. The expected goals of the experiment in this article should follow the requirements of the project. The time requirement for data storage is to insert more than 35,000 data per second, that is, the insertion speed is above 20 m/s. The response time requirement for data retrieval is that the average access load time of 100,000 data records is within 1 s.

3.2 Data Preprocessing

Data integrity: The SEMSS scheme can protect data privacy. For all polynomial time attacks, the retrieved encrypted data and related encryption keys will not reveal any information about related keywords. The hash function H is anti-collision, and is negligible for any polynomial time probability algorithm A , as shown in formula 1:

$$\Pr[(x,y) = A(l_k, H) : x \neq y \wedge H(x) = H(y)] \quad (1)$$

RBT is the optimal strategy for the RDT structure, because RBT contains the optimal (and also the largest) number of RAT tree structures. If the blocks d_i and d_j are related or are often used by MapReduceJob, d_i and d_j should be stored in physically close racks, as shown in Eq. 2:

$$dis(T_i, T_j) \propto \frac{1}{sim(d_i, d_j)} \quad (2)$$

4 Investigation and Analysis of Mass Data Retrieval Technology Based on Cloud Computing

4.1 Data Insertion Performance Before and After Secondary Index Based on Cloud Computing

The results of the data insertion performance before and after the establishment of the secondary index in HBase are shown in Table 1. Comparative analysis: the abscissa in the figure is the amount of data, the statistics are in GB, the ordinate is the time, and the unit is seconds. It can be seen from the figure that in the case of 4 nodes inserting less than 20 GB of data in parallel, if HBase does not build an index, the average insertion speed can reach about 155 m/s, and the average insertion speed of each node can reach 30 m/s.

Table 1. Data insertion performance before and after the establishment of secondary index in HBase

The amount of data	Indexed HBase	HBase not indexed
10G	49	55
20G	112	120
30G	190	208
40G	270	289

If the index is established, the average insertion speed is about 150 m/s, and the average single-node insertion speed is 28 m/s due to Observer's interception, index

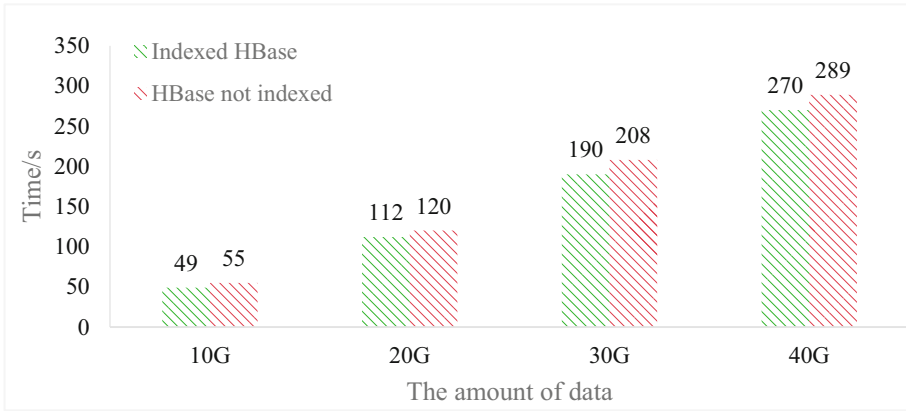


Fig. 1. Data insertion performance before and after the establishment of a secondary index in HBase

creation, and more time to insert index data into the table. That is to say, indexing makes each node The average insertion speed is 2 m/s slow as shown in Fig. 1. With the amount of data within 20 GB, whether HBase is indexed or not, the insertion speed is basically in a linear growth relationship with the amount of data. But at 40G data volume, because the total amount of inserted data exceeds the parallel processing capacity of the cluster, the data will be blocked, and the average processing time will increase by about 10%. According to the project’s expected goal, the data insertion speed can meet above 20 m/s regardless of whether the index is established.

4.2 Data Integrity Verification Based on Cloud Computing Massive Data Retrieval Technology

The authorized user obtains data according to the input keywords in the keyword sets of different sizes, and this process only requires an addition operation. Figure 2 shows the time consumption of generating different sets of keywords. The SEMSS scheme is more efficient due to the encryption operation used in the calculation of the TRSE scheme.

Through the analysis of the query results, the returned top-k documents are more in line with the user’s search intent, even if the query keywords submitted by the authorized user but their semantically related documents will be returned to the user, improving the recall rate. For example, if the user submits the keyword “network”, the keywords of “internet”, “meshwork” and other similar meanings are obtained through semantic mining and used for retrieval. Those documents that contain multiple keywords at the same time also mean that they are more in line with the user’s query need. The search and sort results will also be higher. The experimental results show that the recall rate of the SEMSS scheme is higher than that of the TRSE scheme. The higher the semantic measure of the keyword, the more comprehensive the retrieved match.

Different scales of expanded keyword sets have little difference in the retrieval time of matching documents. Therefore, the retrieval time of the scheme is not affected by

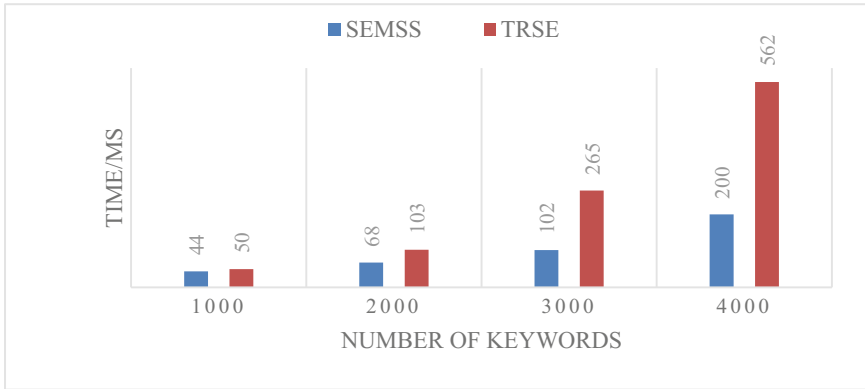


Fig. 2. Time consumption of generating data for different numbers of keywords

the function of semantic expansion, and the document retrieval rate is guaranteed, so it can be verified that the scheme is feasible.

5 Conclusions

The performance of big data storage and retrieval has been faced with the work of optimization and improvement in the face of the explosive growth of data volume. Based on the existing research, this article summarizes the existing results and experience, deeply analyzes the specific process of ciphertext retrieval, breaks the routine use of traditional TF-IDF calculation methods for the preprocessing of massive documents; builds a cloud computing environment, and establishes a key The content retrieval index model of words. In the research of searching specific functions, the cloud server obtains the relevant top-k sorted ciphertext set by calculating the similarity value, and conducts in-depth research on semantic expansion. And collect data simulation experiments to verify that the proposed scheme can obtain better search accuracy and successfully retrieve the data integrity of the technology.

References

1. Xu, H.: Research on mass monitoring data retrieval technology based on HBase. *J. Phys. Conf. Ser.* **1871**(1), 012133 (2021)
2. Koniyo, M.H., Sudarma, M.: Data retrieval service on poverty blockchain-based on regional government. *Bull. Soc. Inform. Theory Appl.* **4**(1), 6–11 (2020)
3. Kironde, H., Morris, B.D., Goel, A., et al.: Retriever: data retrieval tool. *J. Open Source Softw.* **2**(19), 451 (2017)
4. Raad, N., Hasan, T.M., Shakir, A.C.: ECC based data retrieval using LoRaWAN technology. *Int. J. Eng. Technol.* **7**(4), 4918 (2018)
5. Yizi, W.U., Wei, W., Yang, H., et al.: Gas chromatography-mass spectrometry data analysis algorithm based on sparse model and its application in resolving severely overlapped peaks. *Chin. J. Chromatogr.* **36**(8), 772 (2018)

6. Zhang, S.: Large-scale ship fault data retrieval algorithm supporting complex query in cloud computing. *J. Coastal Res.* **97**(sp1), 236 (2019)
7. Wei, Y., Xin, G., Wang, W., et al.: Personal data retrieval and disambiguation in web person search. *IEICE Trans. Inf. Syst.* **E102.D**(2), 392–395 (2019)
8. Ahmad, S., Zahra, S.R.: SDRED: smart data retrieval engine for databases. *Int. J. Inf. Technol. Comput. Sci.* **10**(6), 1–10 (2018)
9. Lucas, D., Antoine, R., Yves, A., et al.: Monitoring glacier albedo as a proxy to derive summer and annual surface mass balances from optical remote-sensing data. *Cryosphere* **12**(1), 1–37 (2018)
10. Gregory, K., Groth, P., Cousijn, H., et al.: Searching data: a review of observational data retrieval practices in selected disciplines. *J. Am. Soc. Inform. Sci. Technol.* **70**(5), 419–432 (2019)
11. Swami, R., Das, P.: An effective secure data retrieval approach using trust evaluation: HBSEE-CBC. *Int. J. Inf. Commun. Technol.* **17**(4), 403 (2020)
12. Wang, F.: Research on construction of road network database based on video retrieval technology. In: *MATEC Web of Conferences*, vol. 139, no. 6, p. 00110 (2017)



Data Analysis of Pharmacy Training Management System Based on Apriori Algorithm

Hui Zheng¹(✉), Dapeng Liu¹, Jing Zhou², Teng Ma¹, and Yan Wang¹

¹ Zaozhuang Vocational College of Science and Technology, Tengzhou 277500, Shandong, China

² Tengzhou Secondary Vocational Education Center School, Tengzhou 277500, Shandong, China

Abstract. As a mature industry management standard, ISO quality management system has undoubtedly become a good choice for hospitals to change management mode. At the same time, it also enriches the connotation and extension of the quality management system in practice, so as to continuously adapt to the characteristics of hospital development and the requirements of the development of medical and health industry. This paper analyzes the Apriori algorithm and data mining, summarizes the ISO quality management system and pharmaceutical professional knowledge, discusses the concept and composition of ISO quality management system and training management system, and studies the quality target bed utilization rate of a people's hospital. The results show that in recent years, the people's Hospital of a City has taken the opportunity of accepting the evaluation of tertiary hospitals, paid attention to quality and sought development, Accelerate project construction and expand wards on a large scale. Especially in the past three years, the utilization rate of hospital beds has increased by leaps and bounds.

Keywords: Apriori algorithm · Pharmaceutical expertise · Training management system · ISO

1 Introduction

Association rule mining is an important part of data mining research. It is the most classic of many data mining knowledge types. Among many association rule mining algorithms, Apriori algorithm is the most typical and influential Boolean association rule mining algorithm.

With the continuous development of science and technology, many experts have studied Apriori algorithm. For example, Tian m, Zhang L, Guo P found the association rules of defect data by using Apriori algorithm based on the 7-year defect data of RPDs of State Grid Corporation of China. In detail, AR between different types of PRD was found and analyzed, such as defective parts and defect causes [1]. Tao and Li use the improved Apriori algorithm to mine the association rules of enrollment information and get the

factors affecting enrollment. It is found that there are significant differences in student registration rates under different rules [2]. Maharjan m uses Apriori algorithm to build a roadmap to analyze consumers' online purchase behavior. Through this analysis, online retailers can easily determine the dimensions affecting online shopping and formulate effective marketing strategies [3]. Although the research results of Apriori algorithm are quite rich, there are still deficiencies in the ISO reform of drug training management.

In order to study the ISO reform of drug training management based on Apriori algorithm, this paper studies Apriori algorithm and drug training management, and finds association rule mining. The results show that Apriori algorithm is conducive to the ISO reform of drug training management.

2 Method

2.1 Apriori Algorithm and Data Mining

The mining steps of Apriori algorithm mainly include two links: one is frequency, that is to find all frequent itemsets according to support; The second is strength, that is, the algorithm generates various association rules according to the credibility. If you want to obtain new knowledge or prove previous theories through a large amount of data, you need corresponding data mining methods. At present, the commonly used mining methods include frequency analysis, frequent itemsets, association rules, classification rules and artificial neural networks. They are widely used in traditional Chinese medicine data mining and database mining. If the transaction database is large, the number of candidate itemsets will increase, and the number of candidate itemsets will increase exponentially, which will greatly increase the memory burden. In addition, the database must be scanned every time the frequent itemsets are determined, which takes a lot of time, which makes the algorithm inefficient, and its efficiency will be lower and lower with the increase of the size of the database.

2.2 Standardization of Pharmaceutical Training Management

(1) Introduction to ISO quality management system

ISO series standards summarize the essence of various quality management practices and theoretical research [1]. ISO series quality management system standards have been widely used all over the world, mainly in line with a common demand gradually generated with the promotion of the concept of total quality management, that is, to continuously produce high-quality products, we need to establish an effective quality management system to continuously improve product quality [2]. In the early 1990s, with the accession of ISO family standards based on British BS5750, China set off an ISO family fever. It summarizes the management experience of advanced enterprises in various countries and adapts to the development of global economic integration [3]. Through the evaluation and supervision of ISO quality management system, it can prove whether the organization's quality management meets the requirements and confirm the organization's quality management ability [4]. ISO quality management system has a unified and standardized

implementation standard. When the organization is put into use, it does not need to make large-scale modifications, and can have more energy to pay attention to and meet the needs of external customers [5].

(2) Composition of ISO quality management system

Typical quality management system documents are generally divided into three levels: quality manual, quality management system procedure documents, operation documents and work records [6]. The preparation of system documents shall meet the requirements of systematicness, compliance and coordination. Structured method is used to build document system. Structured document system is characterized by the order and relevance of documents. As the highest level management system document, the management manual can refer to the procedure documents to achieve the clarity of relevant management activities [7]. The procedure documents specify the details of activities, and various types of operation documents can be referred to when necessary. Operation documents include documents prepared for the effective operation of the system, documents that can be directly used for work guidance, appropriate external documents, etc., and specify the filling, archiving and preservation methods of records [8]. Work records include: records and forms related to products, services and related auxiliary activities. During the management and service process, corresponding work records must be established to confirm the legitimacy of the management and service process and results [9]. The international organization for Standardization ISO has standardized the contents and requirements of total quality management and promulgated ISO series standards. ISO quality management system is a unique management mode suitable for any organization based on the theory of total quality management. ISO quality management system is the concrete embodiment of total quality management thought, the application of total quality management practice method, and the theoretical basis of ISO standard.

(3) Concept and composition of training management system

The training management system is not a simple multiple training, nor is it a mechanical or single dimension training management. This is a process of top-level design, resource integration and effective implementation of training management, with plans, methods and objectives, so that employees can be better qualified for the current position and prepare for higher-level positions. The middle layer is the training management resource layer, which is the deployable or available resources provided by enterprises for the effective implementation of training strategies and systems. The top level of training operation is a series of key behaviors taken by enterprises in the process of implementing training strategy and using various training resources to ensure effective and orderly training. The relationship between the three levels of training management system is gradual, interactive and common

development. A perfect training management system is not only a necessary guarantee to ensure the integrity of talent training and training system, but also an important guarantee to ensure the continuous and effective operation of training [10].

(4) Integration of ISO14001, ISO9001 and safety standardization

Requirements with their own differences cannot be forced to integrate. Forced integration will not achieve good results, but will weaken the strength of document requirements. Therefore, the establishment of system management documents should be handled according to overall management and local management [7]. The integration of documents depends on the scale of the organization, the characteristics and types of products or activities, the process of products or services and the complexity of their interaction, the quality and ability of management and employees. ISO quality management system can define the quality objectives of each department and continuously monitor the realization of objectives to ensure their realization; Timely revise and supplement documents to achieve continuous improvement; Define the work scope of each department, responsibilities and authorities of each person, and complete various works and tasks with high quality through the effective operation of system process methods [11]. Hospitals can learn from the more systematic management methods in ISO quality management system to build a comprehensive administrative management system.

2.3 Association Rule Mining

The association rules of data are the interdependence between data, e). Support is the percentage of transactions that contain both a and B in R, also known as probability; Confidence is the percentage of B when the thing in R already contains a, which is called conditional probability. Support, confidence and promotion are defined as follows (1–2):

$$\text{sup}(A) = P(A) \quad (1)$$

$$\text{conf}(R) = P(B|A) \quad (2)$$

A and B are called the antecedent and the consequent of association rules, respectively.

Support is expressed as the ratio of the trading volume of C and D to the total trading volume, and confidence is expressed as the ratio of the trading volume of C and D to the trading volume of C. The formula of support and confidence is described as follows (3):

$$\text{support}(C \Rightarrow D) = |CUD|/|E| \quad (3)$$

There are limits to support and confidence. Strong rules are rules that satisfy both support and confidence thresholds, and frequent itemsets are itemsets that meet support thresholds.

3 Experience

3.1 Object Extraction

This study intends to use SPSS software to improve the classical Apriori algorithm to realize classification association mining. The parameter setting is the same as that of association rule mining. The classification rules here are called classification association rules. Its essence comes from association rules, but it is different from association rules in the general sense. Its characteristic is to further limit the content of antecedents and antecedents in association rules by setting conditions. Classification association mining mainly constrains the antecedent *a* of the above association rules into a set of conditional attributes, and selects *B* as classification attribute or decision attribute.

3.2 Experimental Analysis

Firstly, the transaction data sheet is established in Microsoft Excel software. After checking again, make perspective views respectively, and convert the perspective views into Excel data tables to perform association rules and cluster analysis operations; Then, the association rules and network system learning program in IBM SPSS modeler 18.0 software package are used to learn the a priori operation of association rules and network system modeling, as well as the preliminary network diagram, including target and mark link degree, and the relevant network diagram and calculation result data table are constructed, including confidence and support; Calculate the link degree in the network diagram and draw the network diagram under different link degrees; Select projects with high confidence and support and re create Excel data sheet; IBM SPSS statistics 22.0 was used for cluster analysis, Pearson correlation calculation method was used to calculate the original data, and a tree model including target, frequency and relative aggregation distance was constructed. Then, the statistical objects contained in the clustering items with different relative distances are counted, the statistical data are and the graphics are drawn; After calculation, modeling and statistics, the results are presented in the form of graphical model; Finally, the chart features are described in the professional language of traditional Chinese medicine, and the relevant laws are analyzed and summarized.

4 Discussion

4.1 Quality Objective Control and Outstanding Effect of Continuous Improvement

According to ISO quality management standard, a municipal people's hospital has established a quality objective system to realize objective management. Each professional department shall formulate corresponding specific decomposition quality objectives according to the medical service scope and supporting work scope of the undergraduate department. After nearly three years of operation, the continuous improvement effect of the quality objectives of the people's Hospital of city a is obvious, which has become a powerful evidence for the continuous improvement of the medical quality of traditional Chinese medicine hospitals in the grade evaluation of general hospitals. See Table 1 for the change trend of bed utilization rate of people's Hospital of city A.

Table 1. Bed utilization rate of a municipal people’s Hospital in recent 3 years

Particular year	Bed utilization
2019	32%
2020	42%
2021	51%

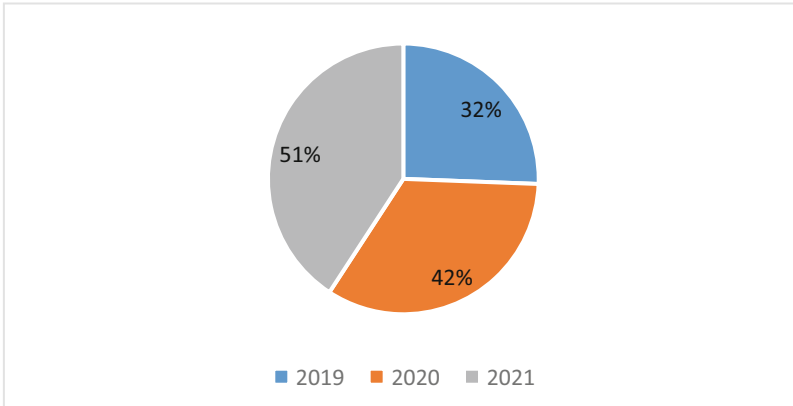


Fig. 1. Bed utilization rate of a municipal people’s Hospital in recent 3 years

It can be seen from the above that the bed utilization rate of the people’s Hospital of city a is 32% in 2019, 42% in 2020 and 51% in 2021. The specific results are shown in Fig. 1.

With the improvement of people’s material and cultural level, people’s demand for health is also increasing, and the phenomenon of long illness in the past has improved. In addition, in recent years, the people’s Hospital of city a has taken the opportunity of accepting the evaluation of tertiary hospitals to grasp quality, seek development, speed up project construction and expand wards on a large scale. Especially in the past three years, the utilization rate of hospital beds has increased by leaps and bounds.

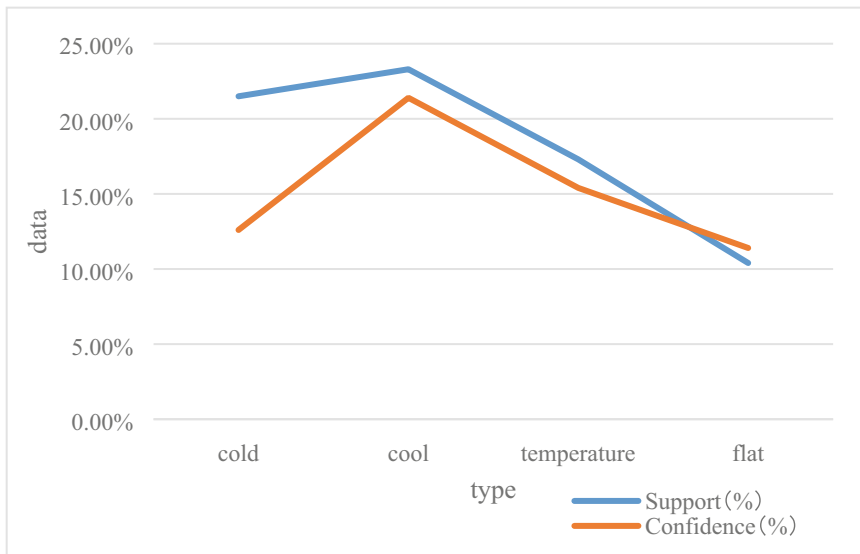
4.2 Correlation Analysis Between Drug Properties and Efficacy

During the association analysis of three, four and five data, in order to more intuitively see the association relationship between the data, this topic plans to use SPSS software to set the corresponding parameters and build a point diagram to display the association rules of three and four data from an intuitive point of view. The correlation analysis of four Qi efficacy is shown in Table 2.

Table 2. Four Qi efficacy correlation analysis

	Support (%)	Confidence (%)
Cold	21.5%	12.6%
Cool	23.3%	21.4%
Temperature	17.3%	15.4%
Flat	10.4%	11.4%

It can be seen from the above that the support of cold air is 21.5%, the confidence is 12.6%, the support of cool air is 23.3%, the confidence is 21.4%, the support of warm air is 17.3%, the confidence is 15.4%, the support of normal air is 10.4%, and the confidence is 11.4%. The specific presentation results are shown in Fig. 2.

**Fig. 2.** Four Qi efficacy correlation analysis

It can be seen from the above that the best laws excavated include: clearing cold and detoxification, clearing heat and detoxification, relieving pain and swelling, cooling and detoxification, warming wind and dehumidification, relieving pain and calming wind and dehumidification. Specifically, it has the highest probability of cold attribute and heat clearing and detoxification attribute at the same time.

5 Conclusion

At present, the world economic situation is undergoing unprecedented changes. Since several industrial revolutions, the economy and aquatic products of all countries in the

world have improved significantly. In the past feudal period, China also moved from closed door to planned economy, and now it has entered the form of market economy. People's domestic water products have been improved, they pay more attention to health than before, and the development of medical undertakings and drug trade has also accelerated. The results showed that the probability of cold attribute and heat clearing and detoxification attribute was the highest.

Acknowledgement. 2021 Zaozhuang City Federation of Social Sciences Applied Research Project: Construction and Application of Comprehensive Skills Training Courses for Pharmacy Based on Professional Ability Training (No. LX2021195).

References

1. Tian, M., Zhang, L., Guo, P., et al.: Data dependence analysis for defects data of relay protection devices based on apriori algorithm. *IEEE Access* **8**, 120647–120653 (2020)
2. Li, T.: Mining association rules on enrollment information of higher vocational colleges using the apriori algorithm. *J. Adv. Comput. Intell. Inform.* **23**(4), 775–781 (2019)
3. Maharjan, M.: Analysis of consumer data on black friday sales using apriori algorithm. *SCITECH Nepal* **14**(1), 17–21 (2019)
4. Wulansari, Z.: Application of data mining to determine product administration using apriori algorithm. *JOSAR (J. Students Acad. Res.)* **4**(1), 18–32 (2019)
5. Ariesty, W.W., Supriyatin, W., Astuti, I.: Marketing strategy for the determination of staple consumer products using FP-growth and apriori algorithm. *Jurnal Ilmiah Ekonomi Bisnis* **24**(3), 225–235 (2019)
6. Marimuthu, M.: An enhancement of apriori algorithm for shopping cart analysis of customers. *Int. J. Adv. Sci. Technol.* **29**(8), 2124–2131 (2020)
7. Zhang, Z., Chai, N., Yuan, L., et al.: Base types selection of product service system based on apriori algorithm and knowledge-based artificial neural network. *IET Collab. Intell. Manufact.* **1**(2), 29–38 (2019)
8. Jantan, H.: Association rule mining based crime analysis using apriori algorithm. *Int. J. Adv. Trends Comput. Sci. Eng.* **8**(1.5), 18–24 (2019)
9. Gao, P., Wang, J., Liu, P.: Research and improvement of apriori algorithm based on hadoop. *Int. J. Adv. Netw. Monit. Controls* **3**(3), 100–105 (2019)
10. Vadivazhagan, K., Karthikeyan, M.: Mining frequent link sets from web log using apriori algorithm. *J. Comput. Theor. Nanosci.* **16**(4), 1395–1401 (2019)
11. Yu, H.: Online teaching quality evaluation based on emotion recognition and improved AprioriTid algorithm. *J. Intell. Fuzzy Syst.* **40**(5), 1–11 (2020)



Health Qigong Sports Network Learning Platform in Cloud Computing

Zhiguo Tian and Wei Diao^(✉)

Zhangzhou Health Vocational College, Zhangzhou 363000, Fujian, China

Abstract. Cloud computing is a network-based computing technology that processes and analyzes data through the Internet and provides the results to users. At the same time, in the context of the rapid development of the Internet, online learning platforms have become a popular trend, and their development prospects are very broad. At the same time, as people's living standards continue to improve, people pay more attention to their physical fitness and physical exercise. Cloud computing, as a key technology for the construction of a Health Qigong (HQG) sports network learning platform, plays an important role in the learning and dissemination of HQG. The data in the HQG network learning platform is mainly realized through cloud computing technology. It can not only provide all the data resources needed to process a virtual community based on the Internet environment. This article uses questionnaire surveys and data analysis methods to explore HQG and sports network learning platforms based on cloud computing and related knowledge, and provide technical support for personalized services for HQG learners. According to the survey results, the HQG sports network learning platform fully cooperates with the learning mode of HQG online learning, caters to the tastes and needs of contemporary people, and has achieved better results in physical fitness, cultivating sentiments, weight loss, and stress relief.

Keywords: Cloud computing · Health Qigong · Online learning platform · Cloud platform

1 Introduction

HQG is one of the representatives of the traditional sports of the Chinese nation and an important part of the long-standing Chinese culture. It has been continuously developed and improved in a long history, has a good training effect on fitness, can be seen everywhere in people's daily life, and is also widely used in university physical education today. Cloud computing is a new type of distributed computing model based on multiple technologies. It provides users with resources through Internet technology. Data and information can be exchanged in this network. In this process, the original database must be divided into user permissions. When the server uploads the user to the cloud and stores it in the terminal device, the corresponding information service will be generated. Its appearance caters to the needs of online learning innovation in the education field. Therefore, based on cloud computing, the establishment of the HQG sports network

learning platform meets the needs of the development of the times and people’s health needs.

From the current point of view, the research results on cloud computing and HQG teaching are very rich. For example, Mingtong proposed that based on cloud computing, a road generation algorithm based on Voronoi diagram is proposed, which can effectively make up for the shortage of road changes generated by the L system [1]. Zheng Shaoxian believes that the perfect combination of physical education and informatization “cloud space” is conducive to breaking the constraints of space and time on learning, and provides conditions for the further development of physical education management [2]. Lu Senlin pointed out that carrying out the teaching of HQG, a traditional national sports project, and incorporating modern elements into it will help actively intervene in the mental health of college students [3]. Therefore, this article combines the relevant knowledge of cloud computing to explore the design of the HQG sports network learning platform, which meets the requirements of the modern physical education model, and has important research significance and application value.

This article mainly discusses these aspects. First, the content of cloud computing and related knowledge is explained. Then, it elaborates on HQG and the network learning platform. In addition, it also introduces the application of cloud computing virtual machine allocation algorithm in the health qigong sports network learning platform. Finally, a questionnaire survey was carried out around the HQG sports network learning platform, and the survey results and analysis conclusions were drawn.

2 Design of Health Qigong Sports Network Learning Platform in Cloud Computing

2.1 Cloud Computing

Cloud computing is a new network service model based on a distributed storage environment. It integrates a large amount of data to form a large and complex virtualized system. It uses Internet technology to operate multiple physical devices, such as servers and storage facilities, in a virtual environment. It is a means of providing information resources developed on the basis of traditional computing technology, with high efficiency and stable performance. It can also effectively integrate dynamic and heterogeneous information resources on Internet services [4, 5]. The hardware and software components of cloud computing are shown in Fig. 1.

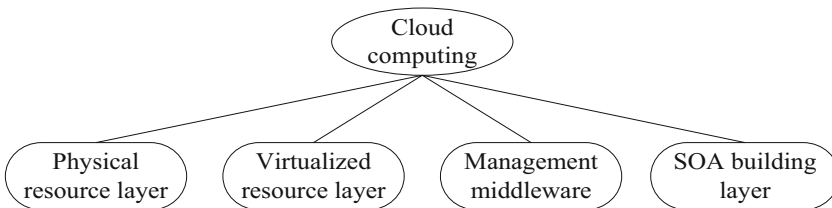


Fig. 1. Composition of cloud computing

Generally speaking, cloud computing has these characteristics. One is virtualization. The object of cloud computing is to store a large amount of data on the Internet, and users can communicate through the Internet. Therefore, the higher the degree of virtualization, the better the system performance. The second is real-time. With the passage of time, the ability of cloud computing to process data is also continuously improving, so there are higher requirements for storage space. The cloud computing platform can quickly, efficiently and accurately provide services to customers according to user needs, and at the same time can ensure the stability of the system during operation through background management. The third is stability and reliability. The core of the module is a stable and reliable server to ensure its security performance. At the same time, it is also necessary to ensure that the background data will not cause system crashes or paralysis due to hacker attacks or malicious program damage. The fourth is flexible service. The specific configuration of the resource pool can be flexibly set, added or reduced according to the business load. It can not only avoid the reduction of equipment utilization caused by the over-allocation of resources, but also solve the inconvenience of adding computing resources to the rising business demand. The various services provided by the cloud computing platform can be accessed anytime and anywhere through the Internet by users using various terminals (such as desktop computers, notebooks, tablets, mobile phones, etc.) [6, 7].

2.2 HQG and Online Learning Platform

Qigong achieves the purpose of cultivating people's vitality by adjusting posture, breathing, and regulating spirit and energy. In order to get rid of unscientific, non-standard, and non-standard exercise methods for Qigong, and to ensure the development of Qigong in a healthy direction, the term "HQG" comes from this. Qigong health care emphasizes its fitness function. Through long-term training, the public can prevent and treat diseases, improve physical functions, and promote physical health. Such exercise is called HQG.

HQG is one of the representatives of traditional health-preserving culture and has a long history in China. It has been circulating for thousands of years and has benefited from generations of Chinese. Its fitness value has been gradually confirmed by many scientists at home and abroad through empirical research. In the context of market economy, the return of promoting excellent traditional culture has become an inevitable condition for the development of the times. Therefore, studying the inheritance, development and healthy innovation of qigong has important social and human values. In the context of modern society and culture, the popularization and promotion of HQG is consistent with the rise of modern national fitness exercises [8, 9].

Only by adapting to the changes of the times, adapting to the changes of the times, and constantly updating and evolving without losing its traditional characteristics, can a culture become a culture with vitality. Today, the Internet has become an important tool and foundation for modern life and work, and an important platform and medium for spreading sports culture. HQG is a unique national traditional culture in China. At the same time, international sports are developing in different countries, and this trend is intensifying. Therefore, in this context, the spread of Chinese HQG should be based on the development and spread of traditional culture. Therefore, in the Internet age, the

use of online learning platforms has great contemporary and practical significance for the development of HQG and healthy culture.

The design of the HQG network learning platform is mainly aimed at two directions. One is from the user’s point of view. The HQG network control system is divided into three parts: the back-end management module and the front-end maintenance. The background management system is responsible for further processing of data storage and server operation. Its functions include system initialization settings and database operations. The other is related services such as member account registration and password modification [10, 11].

2.3 Application of Cloud Computing Virtual Machine Allocation Algorithm in HQG Sports Network Learning Platform

The data is stored through cloud computing technology, and the sports health qigong network matrix is transformed into multiple functions that can be virtualized, callable, and shared resources. Based on cloud technology, the network learning platform can effectively improve its availability, performance analysis and resource utilization. Virtualization technology unifies the computing resources, network resources and storage resources of the network learning platform to construct a virtual network resource pool. By using virtual network resource management technology, it integrates cloud computing resources and optimizes the learning platform [12]. The allocation of cloud computing virtual machines is a multi-objective clustering problem, that is, a multi-objective combinatorial optimization problem. Regarding the allocation of multi-target virtual machines as a multi-dimensional cluster problem, the available resources (boxes) of each virtual node are a c-dimensional vector, and each dimension represents a certain type of resource (CPU, memory, data transmission). The resource (object) of each virtual machine is also a c-dimensional vector. The calculation method is as follows.

$$e_{QW} = \min \sum_k D_k \tag{1}$$

$$e_{AS} = \min \frac{\sum_d G\psi_u}{d} \tag{2}$$

$$G\psi_u = \frac{\sum_b (O_{uk} - \overline{O_u})}{b} \tag{3}$$

Among them, eQW is the number of virtual nodes occupied, eAS is the variance of the server cluster balance load, c is the total dimensionality, is the u-dimensional variance, $G\psi_u$ is the average value of the u-th dimension performance characteristics in all virtual nodes, and O_{uk} is the virtual node k of the u-th dimension performance characteristic value, b represents the number of virtual nodes.

3 Questionnaire Survey Research on the Health Qigong Sports Network Learning Platform

3.1 Questionnaire Design Process

The questionnaire survey selected 80 citizens of T city. Through the issuance of online questionnaires or paper questionnaires, the collection and quantitative analysis of the information filled in by users are carried out to draw conclusions of the questionnaire.

- (1) In the preliminary preparation of the questionnaire, the number of questions should be as concise as possible to avoid fatigue of the interviewees.
- (2) The questionnaire is released. Questionnaires were distributed through online questionnaires, on-site questionnaires, and inviting friends to help ask friends and students around them. A total of 80 questionnaires were distributed, 80 valid questionnaires were returned, and the questionnaire recovery rate was 100%.
- (3) Questionnaire analysis. Organize the collected questionnaire information to get the required information data. Analyze the results of the questionnaire. The analysis results include the effect of the respondents on learning HQG on the sports network learning platform. Some of the results obtained from the questionnaire are as follows.

3.2 Questionnaire Survey Content

The first part is selected 80 citizens of T city, focusing on the learning effect of learning HQG on the sports network learning platform, and investigating the aspects of physical fitness, cultivating sentiment, weight loss, and stress relief.

The second part is to sort out the information collected in the questionnaire to understand the learning effect of 80 citizens on the learning of HQG on the sports network learning platform. Part of the questionnaire survey results are as follows.

4 Questionnaire Survey on the Health Qigong Sports Network Learning Platform

This questionnaire survey focuses on the learning effect of learning HQG on the sports network learning platform, and investigates the aspects of physical fitness, cultivating sentiment, weight loss, and stress relief. The survey results are shown in Table 1.

It can be seen from Fig. 2 that a total of 31 citizens surveyed believe that learning HQG on the sports network learning platform is effective in strengthening physical fitness, cultivating sentiments, losing weight, and relieving stress, while only 8 people believe that it is ineffective. From this, it can be seen that the HQG sports network learning platform fully cooperates with the learning mode of HQG online learning, caters to the tastes and needs of contemporary people, and has achieved good results in strengthening physical fitness, cultivating sentiments, losing weight, and relieving stress.

Table 1. Respondents' evaluation on the learning effect of the HQG online platform

Project	Very effective	Effective	General	Not effective
Physical fitness	8	9	2	1
Cultivating sentiment	7	8	3	2
Losing weight	7	7	4	5
Relieving stress	9	8	3	0

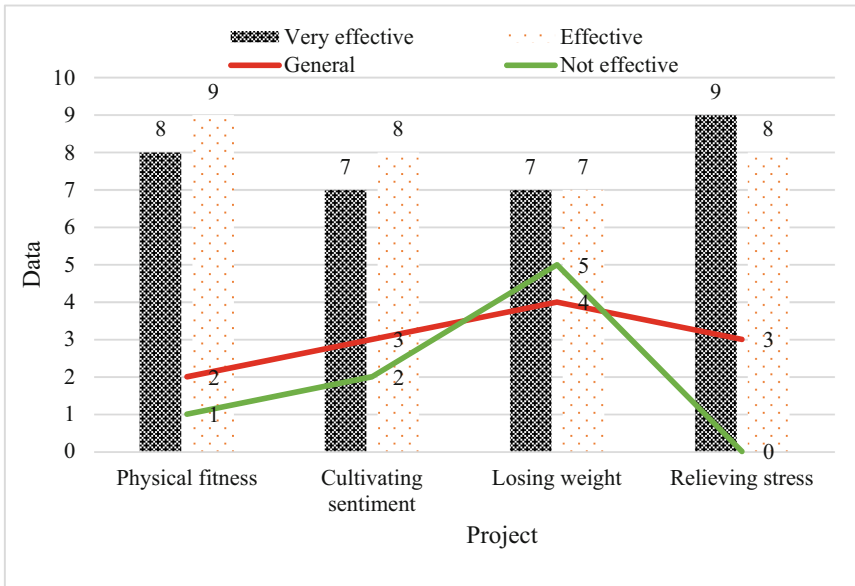


Fig. 2. Respondents' evaluation on the learning effect of the HQG online platform

5 Conclusion

HQG is a traditional national sport and an important part of the long-standing culture of the Chinese nation. It has been loved and studied by the majority of people. At the same time, with the continuous advancement of the Internet industry and cloud computing technology, traditional media and emerging network platforms have entered the digital age. In this context, the HQG network learning platform came into being, which relies on the network operating system to build. On this platform, HQG enthusiasts from all over the country can use it to learn knowledge, transfer and share information, and it has the characteristics of strong real-time, convenience and speed. Therefore, this article studies the design of the HQG sports network learning platform in cloud computing, which has great contemporary and practical significance for the development of HQG health culture.

References

1. Li, M.: Research on the mechanism and influence factors of urban style building based on cloud computing logistics information. *Clust. Comput.* **22**(6), 13873–13880 (2018). <https://doi.org/10.1007/s10586-018-2120-4>
2. Zheng, S.: Research on cloud space construction of physical education teaching based on cloud computing. *J. Yanbian Educ. Coll.* **32**, 149(02), 121–123 (2018)
3. Lu, S., Chen, J., Du, S.: The impact of traditional physical fitness Qigong teaching in colleges on the mental health of college students. *J. Jiangxi Sci. Technol. Normal Univ.* **000**(006), 119–122 (2017)
4. Samriya, J.K., Patel, S.C., Khurana, M., et al.: Intelligent SLA-aware VM allocation and energy minimization approach with EPO algorithm for cloud computing environment. *Math. Probl. Eng.* **2021**(6), 1–13 (2021)
5. Da, S., et al.: VM migration algorithm for the balance of energy resource across data centers in cloud computing. *J. China Univ. Posts Telecommun.* **26**(05), 26–36 (2019)
6. Zhao, Y., Ren, M., Jiang, S., Zhu, G., Xiong, H.: An efficient and revocable storage CP-ABE scheme in the cloud computing. *Computing* **101**(8), 1041–1065 (2018). <https://doi.org/10.1007/s00607-018-0637-2>
7. Sreenu, K., Sreelatha, M.: W-Scheduler: whale optimization for task scheduling in cloud computing. *Clust. Comput.* **22**(1), 1087–1098 (2017). <https://doi.org/10.1007/s10586-017-1055-5>
8. Bharathiraja, N., Ravindhar, N.V., Joel John, M., Pradeepa, K., Gopirajan, P.V., Nagu, B.: Predictive analysis of agricultural crops in cloud computing using sensor networks. *Solid State Technol.* **63**(2S), 5461–5477 (2021)
9. Pande, S.K., Panda, S.K., Das, S., et al.: A resource management algorithm for virtual machine migration in vehicular cloud computing. *Comput. Mater. Continua* **67**(2), 2647–2663 (2021)
10. Chang, P.S., Knobf, T., Oh, B., et al.: Physical and psychological health outcomes of Qigong exercise in older adults: a systematic review and meta-analysis. *Am. J. Chin. Med.* **47**(02), 301–322 (2019)
11. Guo, H., Qiu, J., Yi, L., et al.: Validation of accelerometers combined with machine learning for estimating energy expenditure of health Qigong Baduanjin: 620. *Med. Sci. Sports Exerc.* **53**(8S), 208 (2021)
12. Cao, W., Qiu, J., Yi, L., et al.: Energy expenditure characteristics of Chinese Health Qigong Baduanjin: 679. *Med. Sci. Sports Exerc.* **53**(8S), 231 (2021)



Construction of Stadium Information Management Platform Based on Cluster Intelligent Algorithm

Jing Yang^(✉)

Wuhan Business University, Wuhan 430050, Hubei, China

Abstract. With the popularization of new information technology, it has gradually penetrated into the construction, operation and management of public sports facilities, providing development opportunities for the construction of public sports facilities with the best environmental protection, energy saving, and management efficiency. This paper studies the stadium information management platform based on the cluster intelligence algorithm, understands the relevant theories of the stadium information management platform based on the literature, and then designs the stadium information management platform based on the cluster intelligence algorithm, and through experiments to test the designed platform, the test results show that the overall evaluation of the platform designed in this paper is good, and the number of people with an evaluation of 3 or more accounted for about 80%.

Keywords: Intelligent algorithm · Stadium · Information management · Management platform

1 Introduction

With the rapid development of new information technologies such as 5G technology, artificial intelligence, and blockchain, the development of intelligence in various fields will gradually create infinite new possibilities for people's lives, work, and learning, and it will gradually penetrate into public sports facilities, the construction, operation and management [1, 2]. In order to solve the problems of existing public sports facilities and empowerment, the construction of smart sports facilities promotes the development of public sports facilities to a smarter and higher level, and promotes the path of sports resource allocation, organization and development [3, 4]. The effective integration and advanced communication of sports information resources in public stadiums will play an important role in accelerating the construction of sports units [5, 6]. At the same time, the development of smart stadiums puts forward certain requirements for the intelligent construction and development of public stadiums. It breaks through the existing dilemma of public stadiums managers and emphasizes the role of government leadership and financial support. Facts have proved that the development requirements of smart venues play an important guiding role in the management innovation of public venues, and are also an important symbol of the major development of public venues in my country [7, 8].

Regarding the research on gymnasium management, in view of my country's current management system, some researchers believe that the management system needs to be reformed from seven aspects and improve the practice of standardized support system reform [9]. Under the dual challenges of unemployment reform and international integration into China's large public stadiums, some researchers have proposed three powers to create a comprehensive framework for management and governance reforms, thereby realizing the management of stadiums, sports events, fitness and leisure businesses and major public stadiums. Innovative thinking, the pursuit of improved management level, and through the establishment of venue management team, integration of venue resources and other measures [10]. Some scholars divide the management of stadiums into three methods: transfer of public institutions to enterprises, separation of government and business, and purchase of national public services. In summary, scholars are generally aware of the view that the public stage has changed from the management model of public institutions to the management model of enterprises. Facts have proved that the differentiated mode of management of public stadiums in my country is mainly caused by the characteristics of the joint development of our country's sports industry and sports industry. Therefore, when choosing the management mode of public stadiums, it is necessary to combine the characteristics of practice and development, choose the appropriate management mode, and create more comprehensive benefits.

This paper studies the stadium information management platform based on the cluster intelligence algorithm, analyzes the problems in the stadium information management and the application of the cluster intelligence algorithm based on the literature, and then analyzes the stadium information based on the cluster intelligence algorithm design the management platform, then test the designed platform, and draw relevant conclusions through the test results.

2 Research on the Information Management Platform of Stadiums

2.1 Problems in the Information Management of Stadiums

(1) Manually set general management settings

In the current business model of sports venues in China, due to the different types, uses and systems of sports venues, the sports venues are mainly invested by the state and the independent business model of public institutions. This management model is currently the most important business model for large and medium-sized national stadiums. Because the stadiums are the most prestigious institutions under the jurisdiction of sports management agencies at all levels in the country, under the direct leadership of the sports management agencies, the stadium management departments will operate independently and have independent accounting.

This mode of operation was adopted at the beginning of reform and opening up, and now there is almost no room for development. The venues of this business model are large venues, mostly located in economically underdeveloped areas with large construction sites.

(2) Low utilization rate in stadiums and gymnasiums, closed use

At present, the service conditions of sports facilities in my country are not sufficient. Affected by objective conditions, the utilization rate of sports facilities

is low and the use time is uneven. Most sports facilities have very small space, simple equipment, and little use of site groups, causing a certain amount of waste of resources and increasing a certain cost burden.

(3) Lack of information system support

In addition, the current stage is unmatched by advanced sports academies in terms of high-tech applications. Advanced sports academies have the latest advanced technologies in TV broadcasting, timing, scoring, synchronized widescreen display, network communications, security surveillance, and surveillance prevention, but there is a gap in the requirements for stadiums. On the other hand, most of the existing sports facilities are artificial energy sources, and the outer doors, windows, exterior walls and other building shells have relatively low intelligence. Due to the large area of the stadium, it consumes more energy than other buildings. The energy cost of the stadium is increasing year by year, and the operating cost is increasing.

2.2 Application of Cluster Intelligence Algorithm

Cluster intelligence algorithms are widely used in TSP and other path selection problems, as well as various real-world multi-objective optimization problems. It improved the actual application process to meet the needs of specific problems. The algorithm improves the robustness, convergence and quality of the algorithm solution, and provides better results when solving practical problems.

2.3 Cluster Intelligence Algorithm

In fact, optimization is very important because resources, time and money are often limited. Many problems encountered in real life cannot be constructed like single-objective problems, so multi-objective optimization is required. Appropriate definition optimization is a technique that requires a detailed understanding and rich experience of the problem. There are many types of optimization problems, the most commonly used nonlinear optimization problems. The explanation is as follows:

$$\min f_m(x), (m = 1, 2, \dots, M) \quad (1)$$

The constraints are:

$$h_j(x) = 0, (j = 1, 2, \dots, j) \quad (2)$$

$$g_k(x) \leq 0, (k = 1, 2, \dots, k) \quad (3)$$

Among them: f_m , h_j and g_k are usually nonlinear functions, even integral or differential equations. f_m is the objective function, and when $M > 1$, it is multi-objective optimization.

3 Construction of Stadium Information Management Platform Based on Cluster Intelligent Algorithm

3.1 Analysis of the Construction of the Stadium Information Management Platform

(1) Clarify the management purpose at the stadium level

When formulating management goals, college sports venues should be guided by relevant national policies and public sports service management theories, and follow the principles of fair enjoyment and coverage of teachers, students and employees to achieve different levels of management. Reflect the management in the process of serving the teachers and students of the competition venues and the service to the teachers and students in the management of the venues, effectively guarantee the smooth development of public service facilities and various sports activities, and meet the basic physical exercises of the school teachers and students demand and the quality of life of teachers and students, it not only provides teachers and students with basic sports and cultural enjoyment, but also provides and guarantees basic sports public products and services.

(2) Clarify the operational positioning and responsibilities of stadiums

Judging from previous research results, the quasi-public nature of stadium facilities is basically established, and stadium facilities have complex business and market characteristics. This is because sports venues are based on the school's corporate mission in terms of their functions and responsibilities. They are all open to teachers, students and staff in the school's spare time, and are appropriately disclosed to the society in accordance with the school's relevant documents. This is the school must complete, and as the biggest to achieve the provisions of the sports function in the public service of the facilities to the limit, the first special task that needs to be carried out in the school sports facilities also includes ensuring the college's undergraduate and graduate school's sports teaching obligations, and at the same time ensuring that the school is responsible for sports teams and teams. The responsibilities of individual trainees' coaches in order to promote the organization of competitions, etc., also include the school's requirements for the safety of student sports and the organization of sports activities, as well as guarantees for the organization of various major social activities. The functionality and start-up use of sports venues include responding to the diverse and all-round physical exercise requirements of school faculty, students and social workers, and appropriately opening them to the community, creating sports spaces for businesses, institutions and community residents near the school, and highlight the community value of school sports facilities and so on.

(3) Try to build a comprehensive management model for stadiums and facilities

This comprehensive management model is mainly aimed at different levels of goals, space supply, consumer groups and consumer demand. It strives to allocate space resources scientifically and rationally, and will not cause excessive and irregular opening of stadiums and stadiums and damage and waste, and will not cause the loss of school assets due to excessive negligence at the school stage, while taking into account the public welfare, and maximize the efficiency and utilization

of the school stage. The design of the integrated management model is as follows (Fig. 1):

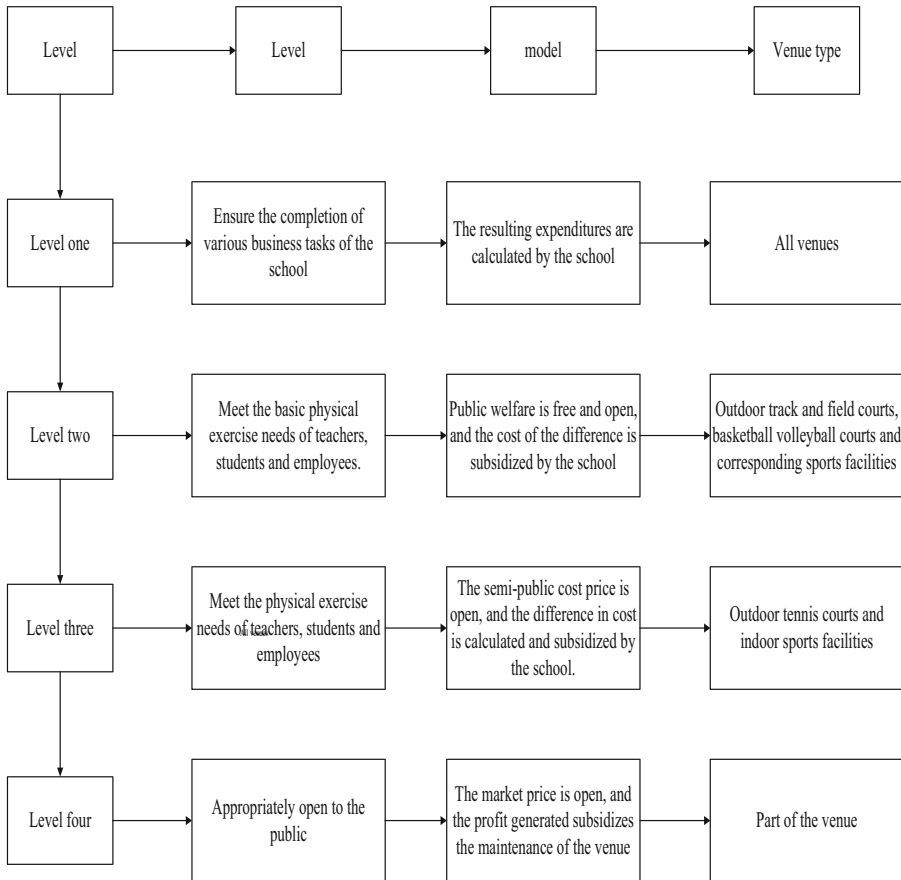


Fig. 1. Integrated management model process

3.2 Functional Design

(1) On-site management module

Campus sports venues should be divided into venue management unit and venue management unit. Each functional module is divided into three parts. The space management part is divided into space use, daily space inspection, and space maintenance. The site management part is divided into site use, daily site inspection and site maintenance.

(2) Equipment management module

1). Sports equipment information management, the main function is to record the basic information of the equipment, including the management personnel who record the use of the equipment, and can query all the basic information of the equipment according to the equipment ID. The equipment manager enters the basic information of the sports equipment purchased by the school, such as serial number, name, product type, manufacturer, contact number, purchase date, location, service life, management personnel, and other sports equipment characteristics. Sort and number devices, set up the system with unique IDs, update device usage and type usage plans, and provide users with basic device information through an easy-to-use interface.

2). The applicant has a scheduled application

Secondly, the equipment reservation/return module mainly works. Users and equipment managers use the equipment reservation/return mode for their respective customers. According to the user ID, the loan and equipment return files will be retrieved in time. Students and teachers apply for the required equipment through the mobile client, and accurately apply for renting the equipment required for class through the sports equipment use plan, and need to book in advance before use.

3). Equipment maintenance module, key functions, inspection, maintenance, repair, and registration of equipment fragments by maintenance personnel. Run daily equipment maintenance files, summarize reports, download all equipment maintenance files and record new and old equipment based on the equipment ID in the platform.

(2) Equipment and facility reservation management module

The main function of space and equipment reservation is to book daily extracurricular sports activities for students and social workers in certain open spaces. This includes three parts: facilities, equipment usage survey, facilities, equipment reservation applications, and facility equipment management.

(3) User rights

The user authentication module is a set of permissions for each management and user structure. Users of different management organizations define different permissions according to their capabilities. The teachers and students of the management organization, as well as the staff of the management organization, have different operation permissions in different management units through account links. Management agencies at all levels build platform content and monitor management cases for different business fields. The main principle of students is to obtain resources and exchange information through mobile clients.

4 Platform Test

Based on the above design concept and design process, we used the platform for a trial. After the practice is completed, a questionnaire survey is issued to the practitioners to collect the practice results. The survey results are shown in Table 1:

Table 1. Platform test results

	Functional evaluation	Operational evaluation	Page review
1	11%	9%	10%
2	11%	12%	13%
3	35%	34%	36%
4	43%	45%	40%

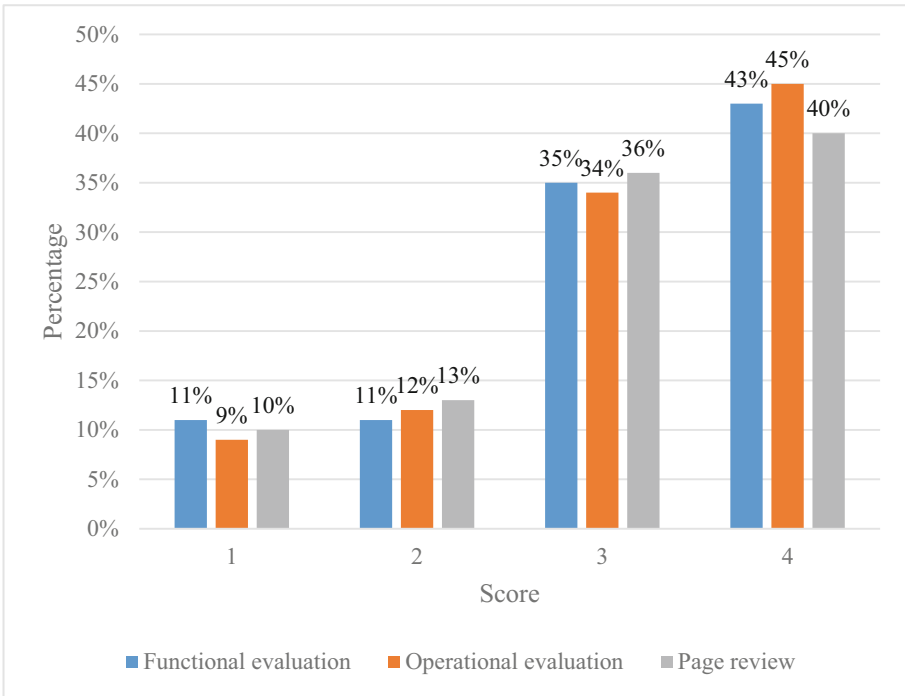


Fig. 2. Platform test results

It can be seen from Fig. 2 that the overall evaluation of the platform is good, and the function evaluation and operation evaluation of the platform are scored above 3 points, which shows that the platform designed this time is more humane.

5 Conclusions

This paper studies the stadium information management platform based on the cluster intelligence algorithm. After understanding the relevant theories, the stadium information management platform based on the cluster intelligence algorithm is designed, and

then the designed platform is tested. The test results are as a result, the overall use evaluation of the platform designed in this paper is relatively good.

Acknowledgements. Wuhan Business School 2020 Annual School-level Scientific Research Project (2020KY016).

References

1. Vaitulionien, R., Augien, D.: Career education in gymnasium: career specialist position. *Baltic J. Career Educ. Manage.* **7**(1), 21–32 (2019)
2. Van Tuyckom, C., et al.: Sports – idrott – esporte – deporte – sportovní The problem of equivalence of meaning in comparative sports research. *Eur. J. Sport Soc.* **8**(1–2), 85–97 (2016)
3. Wang, Y.: Researches on college gymnasium operational management model based on college student's behavioral pattern. *Agro Food Ind. Hi Tech* **28**(1), 3512–3515 (2017)
4. Jia, H.: A study on the predicament and optimization of the opening of gymnasiums in China's colleges to the society. *Agro Food Ind. Hi Tech* **28**(1), 3468–3473 (2017)
5. Chernykh, T.: Problems of human being as a philosophic object in information context of gymnasium social development. *Hum. Bull. Zaporizhzh State Eng. Acad.* **77**, 87–99 (2019)
6. Čorak, L., Milinković, A.: Management of Natural History Collection of Gymnasium Karlovac: preparation for the exhibition in 2016 and plans for the future. *Natura Croatica* **27**(1), 243–255 (2018)
7. Mustafina, G.M., Nikolaeva, N.G., Yermoshin, A.V.: Teaching religion and classical languages in Russian gymnasiums in the middle of XIX - beginning of the XX century: searching for a harmonic correlation. *Man India* **97**(8), 19–28 (2017)
8. Creanor, X.M., Coopoo, Y., Gabriels, G.: Attitudes towards nutritional supplement use amongst adult gymnasium users in Johannesburg North. *South Afr. Sports Med. Assoc.* **29**(1), 1–5 (2017)
9. Tabler, T.: Application of computer-based tools for mathematics teaching at various types of lessons in gymnasiums. *Polonia Univ. Sci. J.* **42**(5), 117–125 (2021)
10. Orevi, I., Ivkovi, L., Jovanovi, S.: Representation ethnological content in geography textbooks for 2nd grade gymnasium. *Zbornik radova - Geografski fakultet Univerziteta u Beogradu* **2017**(65-1a), 41–52 (2017)



Disclosure and Forecast of Stock Issuance Information Based on High Performance Computing and Blockchain Technology

Di Zhao^(✉), Yimei Cao, and Chaoying Xiao

College of Economic and Management,
Shaanxi Fashion on Engineering University, Xi'an 712046, Shaanxi, China

Abstract. Blockchain technology can improve the security and controllability of the blockchain system by controlling membership, but it cannot ensure that all parties will not conspire to tamper with the basic agreement and ultimately harm the interests of other participants. This article established a model: CNN network model based on the distributed mechanism of blockchain technology to predict the stock issuance. The research results show that only when the application cost of blockchain technology is controlled within a certain range, that is, when the BC is 36.01, can the losses caused by the many drawbacks of the traditional model be compensated, thereby promoting the active use of the technology in the securities market.

Keywords: Stock issuance information disclosure · Stock issuance forecast · High-performance computing · Blockchain technology

1 Introduction

Blockchain technology is widely accepted as a disruptive technology, which will greatly improve the accuracy of decision-making by each subject in the securities market and better deal with market fluctuations. Miller proposed that blockchain technology can realize the collaboration of enterprise information at each node, can significantly improve the accuracy of demand forecasting and inventory replenishment capabilities [1]. Morck R pointed out that the stock issuance business stabilizes stock market volatility through reversal trading of high selling and low attracting, and exerts the function of market price discovery [2]. Scheinkman pointed out that the optimism of the amplification of financing transactions and the pessimism of the amplification of securities lending transactions promote market speculation and arbitrage, and improve the efficiency of issuance [3]. Li J et al. built inter-organizational relationships based on the principle of transaction quality [4]. Bris et al. introduced the opportunity to measure the performance of the securities market when the securities market was interrupted [5]. Boulton discussed the horizontal information coordination strategy of stock issuers in a competitive environment [6].

2 Status Quo of the Securities Market and the Blockchain Research Model

2.1 Stock Issuance and Advantages of High-Performance Computing

Stock issuance realizes the price discovery function through short-selling transactions, which can reduce price fluctuations and maintain market stability [7]. By examining the impact of stock issuance on the rate of return and volatility, it can be found that the net financing amount has a positive impact on the rate of return and volatility of the underlying stock [8]. The stock issuance business can promote the speed and degree of response of stock prices to new information, that is, the stock issuance business can significantly improve the issuance efficiency of listed companies [9].

The blockchain assigns witness rights, supervision rights, and decision-making rights equally to each node in the system that actually participates in value exchange activities, and the data records involved in the system are jointly generated and maintained by these nodes [10]. Use the cryptography technology in the blockchain to accurately quantify the value of data information, effectively control data permissions, and ensure user data privacy and rights [11].

2.2 Blockchain Technology and High-Performance Computing Model

The fluctuation value of stock issuance:

$$E = E_w + E_{nb} + E_t - Ic \tag{1}$$

$$E = \frac{\sum_{j=1}^k \sum_{h=1}^k \sum_{t=1}^{n_j} \sum_{r=1}^{n_h} |y_{ij} - y_{hr}|}{2n^2u} \tag{2}$$

In addition, $E = 1$. Thus, the profit function of the stock issuer is:

$$u_h \leq u_j \leq \dots \leq u_k \tag{3}$$

$$E_j = \frac{\frac{1}{2u_j} \sum_{i=1}^{n_j} \sum_{r=1}^{n_j} |y_{ji} - y_{jr}|}{n_j^2} \tag{4}$$

$$E_w = \sum_{j=1}^k G_{jj}p_j s_j \tag{5}$$

Among them, E means not to use blockchain technology [12]. Obtain the optimal profit of stock issuers, sellers and third-party trading companies E and R and the total profit of the overall securities market:

$$E_{jh} = \frac{\sum_{Z=1}^{h_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|}{n_j n_h (u_j + u_h)} \tag{6}$$

$$E_{nb} = \sum_{j=2}^k \sum_{h=1}^{j-1} G_{jh} (p_j s_h + p_h s_j) D_{jh} \tag{7}$$

$$E_t = \sum_{j=2}^k \sum_{h=1}^{j-1} G_{jh}(p_j s_h + p_h s_j) D_{jh} (1 - D_{jh}) \tag{8}$$

Actual fluctuation value of order quantity:

$$D_{jh} = \frac{d_{jh} - P_{jh}}{d_{jh} + P_{jh}} \tag{9}$$

Assuming that the total cost of adopting the blockchain is BC [13]. The cost saved by each node enterprise in the securities market in a single search for a partner is:

$$d_{jh} = \int_0^\infty dF_j(y) \int_0^y (y - x) dF_h(x) \tag{10}$$

$$d_{jh} = \int_0^\infty dF_h(y) \int_0^y (y - x) dF_j(x) \tag{11}$$

Inventory risks can be effectively avoided through information coordination, namely:

$$f(x) = \frac{1}{Nh} \sum_{i=1}^N k\left(\frac{X_i - x}{h}\right) \tag{12}$$

$$k(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{13}$$

In this paper, the profit of the two parts is maximized, and the corresponding optimal solution of the circulation is obtained:

$$h_t = \tanh(w_c x_t + u_c(r_t \Theta h_{t-1}) + b_c) \tag{14}$$

$$h_t = z_t \Theta h_{t-1} + (1 - z_t) \Theta h_t \tag{15}$$

h_t indicates the use of blockchain technology [14]. In this way, the profit function of stock issuers is as follows:

$$P = \sigma t = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (FI_{it} - FI_{it})^2}}{FI_{it}} \tag{16}$$

$$u_{(j|i)} = w_{ij} A_i \tag{17}$$

$$s_j = \sum_i c_{ij} u_{(j|i)} \tag{18}$$

3 Information Disclosure and Prediction of Stock Issuance

3.1 Contents and Steps

(1) Model design

A CNN network model based on the diversification mechanism of blockchain technology is established to predict the stock issuance. The model is as follows:

$$\text{Efficiency}_{i,t} = \alpha_0 + \beta * \text{Policy}_{i,t} + \theta * \text{Controls}_{i,t} + u_i + \varepsilon_{i,t} \quad (19)$$

Among them, $\text{efficiency}_{i,t}$ represents the issuance efficiency of the i -th stock issue target in the T period [15]. $\text{Policy}_{i,t}$ is a policy dummy variable. Its value before and after being selected is 0 and 1 respectively. $\text{Controls}_{i,t}$ represents a series of control variables [16]. u_i is the individual effect of the underlying stock, $\varepsilon_{i,t}$ is a random error term.

(2) Metrics

This paper measures the stock issuance efficiency from two dimensions: the first is the stock price synchronicity [17]. The second is the absolute value of the correlation coefficient between the return of individual stock and the return of the lagging one period market [18]. The auxiliary regression equation was as follows

$$r_{i,t} = \alpha + \beta * r_{m,t} + \varepsilon_{i,t} \quad (20)$$

The goodness of fit is transformed into logarithm to obtain the efficiency of stock synchronicity l_i

$$\text{Efficiency}_{i,t} = \ln\left(\frac{R_{i,t}^2}{1 - R_{i,t}^2}\right) \quad (21)$$

(3) Evaluation index

In order to evaluate the prediction performance of ceemdan-agu model for stock price, this paper uses three regression evaluation indexes: root mean square error (RMSE), mean absolute error (MAE) and R-square (R^2) [19].

4 Stock Issuance Information Disclosure and Prediction

4.1 Double Differential Regression Analysis of Information Disclosure

As shown in Fig. 1, due to the strong endogenous selectivity of the expansion target, that is, the issuance efficiency itself is better, the target is more likely to be selected as the object of stock issuance expansion. Before the implementation of the policy, there are many differences in the average issuance efficiency between the control group and the experimental group. However, with the increase of the number of stock issuance

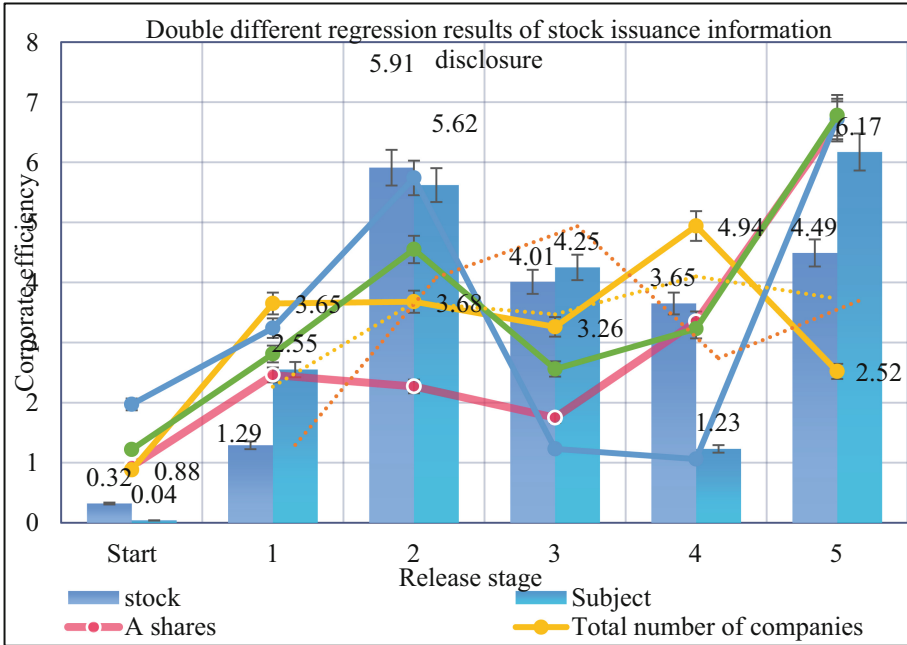


Fig. 1. Double difference regression results of stock issuance information disclosure

Table 1. Regression coefficients obtained by two estimation methods

Num	Stock	Subject	A shares	Total number	Percentage	Multiply
Start	0.32	0.04	0.94	0.88	1.97	1.22
1	1.29	2.55	2.46	3.65	3.24	2.81
2	5.91	5.62	2.27	3.68	5.74	4.55
3	4.01	4.25	1.75	3.26	1.23	2.56
4	3.65	1.23	3.35	4.94	1.06	3.23
5	4.49	6.17	6.68	2.52	6.72	6.78

expansion, the distribution of the issuing efficiency of the selected targets increases, so the grouping randomness is also strengthened.

As shown in Table 1, the regression coefficients obtained by the two estimation methods are consistent, while the coefficient variance obtained by the boot strapped robust standard error estimation is slightly larger than that of the ordinary standard error. The interaction coefficient reflecting the policy effect is significantly negative at the level of 0.1 and 0.01 under the two issuance efficiency indicators, which verifies the conclusion that the implementation of the stock issuance policy does improve the issuance efficiency. The results show that the double difference review has strong robustness.

Table 2. Influence of preference index and transaction cost coefficient

Item	Extremely severe injury	Severe injury	Damage	Blockchain technology	Coefficient of elasticity	Price tradition
Enterprise E	3.26	2.44	2.95	1.36	1.29	2.55
Enterprise R	3.43	4.58	2.84	2.59	5.91	5.62
Trans-service	5.58	5.61	4.71	4.78	4.01	4.25
Preference	3.54	3.6	4.48	2.2	3.65	1.23
Transaction	3.3	2.61	4.67	5.79	4.49	6.17
Scale	5.44	2.1	4.39	3.65	1.23	6.33

As shown in Table 2, with and without blockchain technology, the order quantity of two third-party trading enterprises is affected by preference index and transaction cost coefficient. Compared with the traditional mode, the order quantity under the application of blockchain technology is higher than that under the traditional mode. Blockchain has changed the mode of information ownership, transferring the ownership of information from the state owned by a single owner to the account information shared by all participants in the whole transaction cycle.

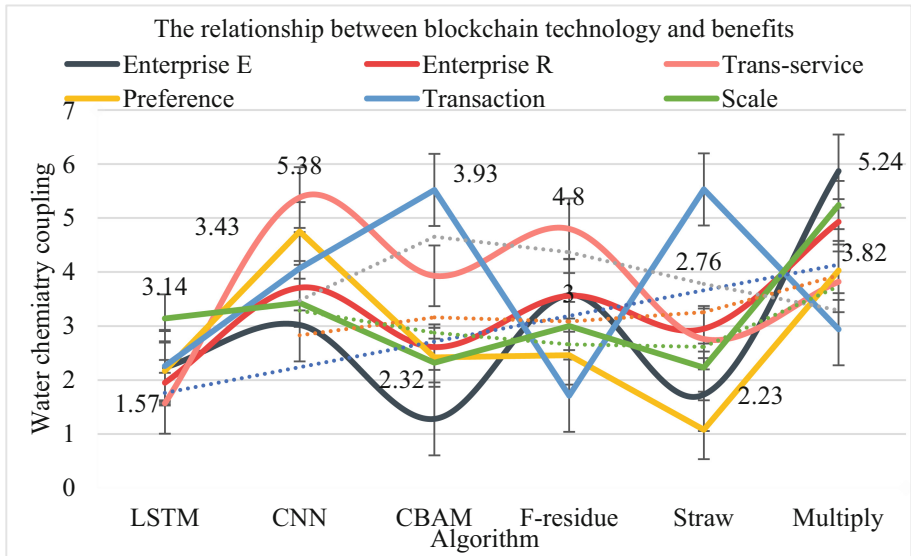


Fig. 2. The relationship between blockchain technology and benefits

As shown in Fig. 2, the results show that the application of blockchain technology may not make the profit of each node enterprise and the overall profit increase significantly. When the application cost of blockchain technology is too high, each node enterprise in the securities market will choose to give up using the technology.

4.2 Implicit Information Disclosure and Numerical Simulation

As shown in Fig. 3, although the LSTM model can fit the general fluctuation of stock price, the fitting accuracy is low, and the prediction effect of CBAM is obviously better than that of LSTM network model and LSTM-CNN network model. The model can predict the stock price reasonably and accurately.

As shown in Table 3, there is a big difference between the predicted value and the actual value of the stock price of the model, which may be caused by the influence of

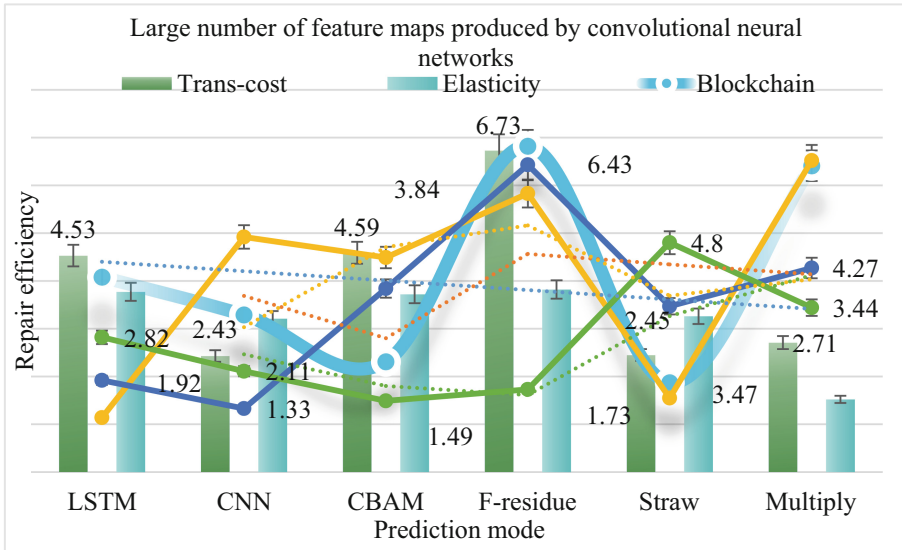


Fig. 3. Large number of feature maps produced by convolutional neural networks

Table 3. The difference between the predicted value of the stock price and the actual value

Item	Trans-cost	Blockchain	Elasticity	Price-trad	Profit	Upgrade
LSTM	4.53	4.08	3.77	1.14	1.92	2.82
CNN	2.43	3.29	3.21	4.92	1.33	2.11
CBAM	4.59	2.31	3.72	4.49	3.84	1.49
F-residue	6.73	6.82	3.82	5.83	6.43	1.73
Straw	2.45	1.87	3.26	1.55	3.47	4.8
Multiply	2.71	6.41	1.52	6.52	4.27	3.44

the government policy or network public opinion on the stock market at that time, rather than the defects of the network proposed in this paper. Therefore, this experiment verifies the effectiveness and feasibility of the proposed network.

5 Conclusions

Because the law does not endow private digital currency with exclusive and compulsory functions, and there is no national capital as the credit basis, blockchain technology does not have the essential attribute of money. However, because it is accepted by the general public in a certain range, it poses a serious challenge to the legal tender system of the country. It is necessary to control the application cost of blockchain technology within a reasonable range, so as to encourage enterprises to actively explore the determination of blockchain technology application.

Acknowledgements. Provincial project: Study on the housing problem of the introduction of talents in Xi'an city under the background of new Settlement.

Project number: 20JK0042.

References

1. Miller, E.M.: Risk, uncertainty, and divergence of opinion. *J. Finance* **7**(32), 1151–1168 (2020)
2. Diamond, D.W., Verrecchia, R.E.: Constraints on short-selling and asset price adjustment to private information. *J. Financ. Econ.* **7**(18), 277–311 (2020)
3. Morck, R., Yeung, B., Yu, W.: The information content of stock markets: why do emerging markets have synchronous stock price movements? *J. Financ. Econ.* **9**(58), 215–260 (2019)
4. Scheinkman, J.A., Xiong, W.: Overconfidence and speculative bubbles. *J. Polit. Econ.* **4**(1), 1183–1220 (2020)
5. Li, J., Myers, S.C.: Around the world: new theory and new tests. *J. Financ. Econ.* **79**(2), 257–292 (2016)
6. Bris, A., Goetzmann, W.N., Zhu, N.: Efficiency and the bear: short sales and markets around the world. *J. Finance* **9**(62), 1029–1079 (2018)
7. Boulton, T.J., Braga-Alves, M.V.: The skinny on the 2008 naked short sale restrictions. *J. Financ. Mark.* **5**(13), 397–421 (2019)
8. Ghysels, E., Santa-Clara, P., Valkanov, R.: There is a risk-return trade-off after all. *J. Financ. Econ.* **76**(3), 509–548 (2020)
9. French, K.R., Schwert, G.W., Stambaugh, R.F.: Expected stock returns and volatility. *J. Financ. Econ.* **19**(1), 3–29 (2020)
10. Guo, H., Whitelaw, R.F.: Uncovering the risk–return relation in the stock market. *J. Finance* **61**(3), 1433–1463 (2016)
11. Salvador, E., Floros, C., Arago, V.: Re-examining the risk–return relationship in Europe: linear or non-linear trade-off? *J. Empir. Finance* **28**(4), 60–77 (2019)
12. Brandt, M.W., Kang, Q.: On the relationship between the conditional mean and volatility of stock returns: a latent VAR approach. *J. Financ. Econ.* **72**(2), 217–257 (2019)
13. Rapach, D.E., Wohar, M.E.: Multi-period portfolio choice and the intertemporal hedging demands for stocks and bonds: international evidence. *J. Int. Money Finance* **28**(3), 427–453 (2019)

14. Glosten, L.R., Jagannathan, R., Runkle, D.E.: On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* **48**(5), 1779–1801 (2020)
15. Turner, C.M., Startz, R., Nelson, C.R.: A Markov model of heteroskedasticity, risk, and learning in the stock market. *J. Financ. Econ.* **25**(1), 3–22 (2020)
16. Black, F., Cox, J.C.: Valuing corporate securities: some effects of bond indenture provisions. *J. Finance* **31**(2), 351–367 (2020)
17. Li, Q., Yang, J., Hsiao, C., et al.: The relationship between stock returns and volatility in international stock markets. *J. Empir. Finance* **12**(5), 650–665 (2019)
18. Hatemi-J, A., Irandoust, M.: The dynamic interaction between volatility and returns in the US stock market using leveraged bootstrap simulations. *Res. Int. Bus. Finance* **25**(3), 329–334 (2019)
19. Smith, L.V., Yamagata, T.: Firm level return–volatility analysis using dynamic panels. *J. Empir. Finance* **18**(5), 847–867 (2019)



Design and Research of Enterprise Tax Accounting Database Under the Background of Big Data

Yuxin Xie^(✉)

Accounting and Taxation, Aba Normal University, Chengdu 610066, China

Abstract. With the development of our country's global economic integration, facing more opportunities and greater challenges, enterprises have higher and higher requirements for the update of management software, and enterprise accounting informatization has achieved considerable development. This article aims to study the design of corporate tax accounting database in the context of big data. On the basis of analyzing the choice of database, the physical and logical structure of Oracle database, and the feasibility analysis of tax accounting management information system, the tax accounting management system and database are analyzed. Finally, the performance of the system is tested. The test results show that the test is successfully passed when 300 users log in. At this time, the average response time of operations after logging in to the system is 2.84 s, which meets the performance requirements of the system in this article.

Keywords: Big data · Tax accounting · Database design · Management system

1 Introduction

Since entering the 21st century, the development of information technology has brought us not only convenience in life, but also convenience in work. With the development of accounting information, traditional manual bookkeeping has been replaced by electronic computer data processing. With the development of tax information, traditional paper tax returns have been replaced by the latest electronic and online tax returns [1, 2]. However, it should be noted that illegal cases such as tax evasion and tax evasion by enterprises using electronic technology are not uncommon. Our country has issued a number of relevant policies to strengthen internal control of enterprises, which have a corrective effect on economy, orderly, double accounting and taxation information [3, 4].

After the creation of database theory, some researchers introduced advanced database system technology, thus providing the basic logical structure of evidence-based accounting information system [5, 6]. Some researchers have clearly combined the technical issues of database systems with the issues of establishing a decentralized orthogonal protection accounting information system. In accordance with the idea of material law, he boldly abandoned the traditional bibliographic bookkeeping principle in the overall design of the computer financial information system, and introduced a hierarchical

model. It introduces tools that can encode various items with financial and non-monetary characteristics. It also explains the general diagrams and data types of items coded through a hierarchical data structure, and creates a data management language for enterprise data management [7, 8]. The concept of the enterprise relational database model provided by some researchers is also used in the field of accounting. In order to infer the relevant models of accounting management and financial accounting, they first introduced the traditional accounting system structure. To analyze and establish a conceptual model of the relational database. Then, further explain how the information is obtained in the database through relational algebra [9, 10]. The entity link diagram created by some researchers creates a highly semantically expressed entity link data model, so that the data model not only describes the data involved in the accounting process, but also emphasizes the company in its data model. There are also researchers who have used both type method and value method in the inventory management process and production environment by guiding decision makers, and have compared and commented on the two types of methods [11, 12]. The research results of predecessors provide a theoretical basis for the research of this article.

On the basis of consulting a large number of related references, this paper combines the choice of database, the physical and logical structure of Oracle database and the feasibility analysis of tax accounting management information system, and designs the tax accounting management system and database, and finally the system. The performance is tested to observe whether the system meets the requirements of this article.

2 Design and Research of Enterprise Tax Accounting Database Under the Background of Big Data

2.1 Database Selection

(1) Access database

As an important factor, Microsoft Access is widely adopted. Access database is also suitable for the development of small database systems. Access can be widely used, mainly due to its strong analytical capabilities and relatively simple operation. However, Access is only provided to a small user system, and cannot provide access services for many application systems at the same time. This is the main flaw of the database system. In addition, this database system is only used on the Windows operating system platform.

(2) Oracle database

Oracle Enterprise has developed an Oracle database management system for actual needs. Compared with other database technology companies, Oracle has the following technical features:

- (1) **Compatibility:** Whether it can pass the standard SQL database query language is an important sign of Oracle's reliability. This also shows that the applications released to users using the Oracle database system can be widely used on SQL databases.

- (2) **Portability:** Due to the large number of Oracle workstations, Oracle is defined as a standard database management system, and Oracle applications can be reused after transplantation, regardless of the size of the hardware configuration.
- (3) **Versatility:** The main advantage of Oracle is that its architecture is more complex. Back-end and front-end editing are key elements of Oracle database application information performance. In addition, you can use this feature to protect multi-user data and process corresponding transactions at the same time.

(3) **SQL Server database**

SQL Server is a popular background management product of Microsoft, especially SQL Server 2005 has become a favorite language of many developers. Compared with other databases, it has many advantages and can be used with many components.

2.2 The Physical Structure and Logical Structure of the Oracle Database

(1) **Physical structure**

Although the Oracle database is a collection of various data types, all methods of actually storing data are in the form of physical documents. The physical structure of the database also represents the format for storing various data on the same physical medium. As for the physical structure of the entire database system, it is mainly determined by the operating system files that make up the database. For each Oracle database, it is mainly composed of physical file systems of the following main data types.

(2) **Logical structure**

Logical database system structure refers to the composition of the database system involved by the user. Explains how the database system stores data logically. Logical storage structures such as table spaces, partitions, and extensions manage the application of physical storage space in the database system. A database system is logically composed of one or many table spaces, each table space is composed of a segment, the segment is composed of regions, and the region is composed of many continuous databases. Of course, some logical structures such as lists and searches are also stored in the tablespace data file.

2.3 Feasibility Analysis of Tax Accounting Management Information System

(1) **Operational feasibility**

In the tax information system research and development process, all operating systems and development environments are developed using the VB6.0 platform. The entire system interface is easier to apply, and the operation of the system is simpler. At the same time, in the VB6.0 environment, the operating system functions are basically the same as those in other computer software. Most of them meet the current tax management habits of enterprises and save a lot of training time.

(2) Economic feasibility

The tax management information system uses a dedicated network, the total number of users is less than the Internet, and the demand for databases and WEB servers is small, so the investment amount is minimal. To use the client, only the most basic computer is required. The support of the B/S model also reduces the maintenance and upgrade costs of the next stage of the tax management information system.

(3) Technical feasibility

At this stage, the technologies used in the development of tax management information systems are mainly mature programming languages and database application technologies. The main techniques used in data analysis are statistics, query and data analysis. The technical requirements are not very high, so it is absolutely technically feasible. At the same time, in the development and discovery stage of the tax management information system, there are already relatively similar systems in the reporting market, combining the characteristics and functional requirements required for the development of the tax management information system. Therefore, after the development of the tax management information system, it will definitely be very useful, The implementation of the entire technology is relatively easy.

(4) Security analysis

In order to ensure the operational safety of the tax management information system and fully guarantee the security of the information contained in the tax management information system, the following three aspects should be considered when designing and developing the tax management information system. First, from the perspective of the users who use the entire system Look, the user login needs to be verified, and can only enter when it is legal. Secondly, for the entire system administrator, the administrator only keeps the relevant modules in the system during the entire system maintenance process, and does not touch the relevant personal data and information of the user. It is safer to analyze the information contained in the tax management information system from the perspective of the administrator. Third, the entire tax management information system is equipped with targeted system abnormality detection functions. If the system has abnormal data processing, illegal user intrusion, abnormal data exchange, etc., the system will issue related warnings, which greatly improves the security of system operation.

3 Experiment

3.1 Identity Verification Registration Module

The system identity confirmation registration module mainly provides login confirmation and registration functions for system users to prevent unauthorized users from entering the system. When logging in, only registered users can access the system. In addition, when different characters enter the system, different interface features will be displayed.

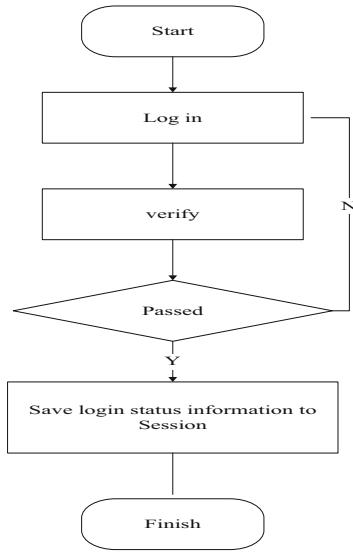


Fig. 1. Flow chart of system login

When registering, after the user submits the registration application form, it will be reviewed by the back-end administrator. Throughout the process, the password uses MD5 encryption technology to ensure system security. The process is shown in Fig. 1.

The logic function used by MD5 encryption technology is shown in the following formula:

$$F(x, y, z) = (x \& y) / ((\sim x) \& z) \quad (1)$$

$$G(x, y, z) = (x \& Z) / (y \& (\sim z)) \quad (2)$$

In the formula, & and ~ respectively represent AND and NOT operations.

3.2 Tax Declaration Module

- (1) Data analysis input function. Since the ERP system has many modules, the financial module contains the tax data that needs to be calculated, so this operation is a method of analyzing and extracting the required information from the massive financial data. The following points may be required to realize this part: based on the data analysis model, Data control technology, the design of this part can refer to the existing ERP system of the tax accounting part.
- (2) Accounting for tax-related businesses. This part is based on the enterprise's application in the independent tax accounting mode of the ERP system, based on the same clearing operations, accounting rules, and data calibers, and automatically charges costs according to the tax accounting formula associated with the imported data. The quarterly relevant tax data of tax-related tax categories of enterprises, the main data comes from the financial data of the ERP system.

- (3) Tax declaration. In this part, after the tax declaration form is issued at the end of the quarter, the relevant data will be automatically obtained according to the results of the tax accounting firm in the previous step, and then a tax declaration form that meets the requirements will be generated. Create a tax department, extract the data in the tax return from the result, and automatically upload it to the tax authority after passing the business self-inspection part. This part of the operation can be done by optimizing the call to the corporate tax reporting system.

3.3 System Management Module

Depending on the role of the user who uses this system, the module permissions that can be used are also different. Therefore, according to the location of the tax service, you need to define their respective usage rights for users in that location. In this way, users who use the same system but different permissions will see different system functions after logging in, so that users can only use system functions within their own permissions, which enhances security.

3.4 System Database Design

In terms of representation methods, there are also many ways to represent conceptual models. The most common representation in the model design process is the entity link method. The so-called method of linking entities is to use E-R diagrams to illustrate the associations between various models. And the following mainly uses E-R diagram to illustrate the structure of the database. (1) The entity is represented by a rectangle. (2) Entity attributes are represented by ellipses. (3) The relationship between the physical object and the entity is represented by a diamond. In this tax accounting database design process, this paper fully considered the internal correlation of the internal data information of each module, and adopted intuitive means to systematically analyze and design the E-R diagram.

4 Discussion

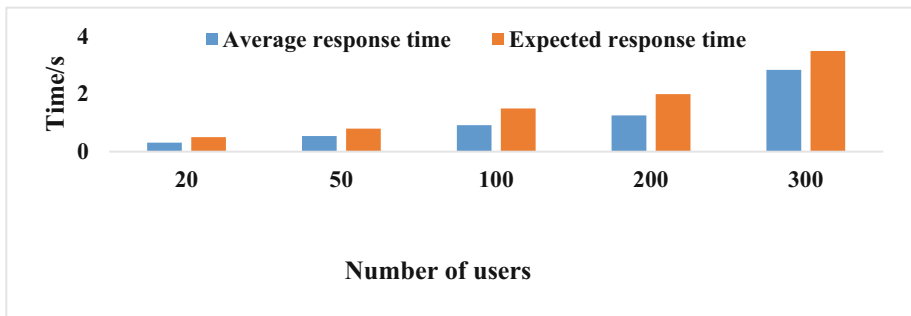
The tax accounting management system has conducted a preliminary performance test before it is put into use. The purpose of the test is to test the response speed and system stability of the system in multi-user situations. Some problems have been found through the test, mainly because the system has some problems when uploading big data files. There may be no response for a long time. It is analyzed that the problem may be caused by the user's computer or network conditions. The specific reason needs to be further studied and analyzed.

In order to grasp the performance of the system under multi-user conditions, the system organized a multi-user online performance test when it was put into trial operation to observe the response time of the system.

It can be seen from Table 1 and Fig. 2 that the average response time during system operation is within 3 s. When 300 users log in, they successfully pass the test. At this time, the average response time of operations after logging in to the system is 2.84 s, which meets the performance requirements of the system in this article.

Table 1. System response time

	Average response time	Expected response time
20	0.31	0.5
50	0.54	0.8
100	0.92	1.5
200	1.26	2
300	2.84	3.5

**Fig. 2.** System response time

5 Conclusions

Under the influence of the continuous development of information technology, tax accounting informatization is a process of continuous development. Information resources are integrated on the basis of data, promote the high integration of information technology and tax accounting, and continue to strengthen the legal protection of tax accounting operations in the information environment. To improve the latest operating mechanism of tax accounting, to meet the actual needs of tax accounting work as the goal, to achieve the rapid development of tax accounting information construction.

References

1. Chen, Y.: Analysis of the effect of computer technology on the path of tax accounting teaching reform in accounting major under the background of China's tax reform. *J. Phys. Conf. Ser.* **1744**(4), 042069 (2021)
2. Melnyk, K.: Implementation of audit functions as a factor of improvement of tax accounting of industrial clusters. *Econ. J. Lesia Ukrainka East. Eur. National Univ.* **3**(23), 177–181 (2020)
3. Zabaznova, D.: Adjustment of the accounting policy of the agricultural holding for reliable tax accounting. *Auditor* **6**(11), 39–44 (2020)
4. Hytis, E.: The informative value of taxes: the case of temporal differences in tax accounting. *J. Acc. Taxation* **11**(8), 130–138 (2019)

5. Odakhovskaya, D., Pechennikova, G.: Accounting and tax accounting of fixed assets, intangible assets and depreciation for development and improvement of approaches to assessment of the company's assets. *Bull. Baikal State Univ.* **29**(1), 113–122 (2019)
6. Nwaorgu, I.A., Abiahu, M., Tapang, A.T., et al.: Deferred tax accounting and financial performance: the listed agricultural firms perspective in Nigeria. *SSRN Electron. J.* **18**(3), 128–143 (2019)
7. Dwitayanti, D.N.: Application of the Permendagri no. 64 2013 on listing of tax accounting hotels. *JOSAR (J. Stud. Acad. Res.)* **3**(2), 77–87 (2018)
8. Abdullaev, A.: Tax accounting and analysis: about scientific research and new approach. *Int. Finance Account.* **2018**(4), 1 (2018)
9. Ibragimovna, K.K.: Accounting of tax benefits. *Asian J. Multidimensional Res. (AJMR)* **10**(5), 26–29 (2021)
10. Cernuca, L.: The accounting and tax regime regarding sponsorship. *CECCAR Bus. Rev.* **2020**(2), 19–29 (2020)
11. Zhou, L., Che, P.: Tax information accounting of marine enterprises. *J. Coastal Res.* **106**(sp1), 272 (2020)
12. Kucharova, I., Pfeiferova, D., Lrinczova, E.: Specifics of cryptocurrencies from an accounting, tax and financial view in a globalized environment. *SHS Web Conf.* **92**(6), 03014 (2021)



Mental Disease Early Warning Analysis Technology Based on Social Media Big Data Algorithm

Jiangnan Li(✉)

School of Education Science, Guangxi Science and Technology Normal University,
Laibin 546199, Guangxi, China
lijiangnan0108@sina.com

Abstract. In today's society, with the rapid economic development, people's living standards have also improved, while at the same time, many psychological problems have been ignored. And mental illness refers to the abnormal reaction of the human body due to various reasons, resulting in some physical and mental disorders and even a series of physiological or pathological changes, such as depression, mental confusion, etc. These unhealthy conditions will bring varying degrees of pain and impact to patients, and at the same time will cause a lot of burden and pressure on the family. Therefore, in the process of medical and health services, it is necessary to timely understand that the patient and his family are facing mental illness. Therefore, how to discover a person's abnormal state of emotion in time, and get timely and effective early warning processing and intervention on it has become a hot topic. This article adopts experimental analysis method and data analysis method, and aims to understand the early warning analysis technology of mental illness through experimental research, so as to effectively and timely monitor and early warning of mental illness. According to the experimental results, when the number of concurrent users is less than 500, the response time of the system is shorter, and the time to get the results of mental health analysis is also shorter, so it can meet the basic needs of mental health analysis and early warning.

Keywords: Social media big data · Mental health · Psychological crisis early warning · Early warning indicators

1 Introduction

With the development of society and the continuous improvement of people's living standards, people are paying more and more attention to their own health problems. However, there are many mental illnesses in real life. For example, anxiety, depression, etc., these symptoms will affect people's normal way of thinking. It has negative consequences for society, affects communication between people and the outside world, reduces work efficiency and causes various mental problems. Therefore, in this case, we must understand our physical condition in a timely and effective manner, and take

corresponding measures to prevent and intervene in order to avoid the deterioration of the disease and endanger human health and safety.

At present, many scholars have conducted research on big data analysis methods and mental health issues. For example, Liang Juan applied big data mining methods to analyze student data information in the college student management system, and found a certain correlation between mental health data and data such as grades [1]. Imran M pointed out that Data analysis can show the value of stock transactions in the economy and indicate a high ability to mobilize capital and diversify risk [2]. Huang Xiaojuan pointed out that although people's material living standards have become higher and higher, there have been a lot of problems on the spiritual level [3]. Therefore, this article starts from a new perspective, combined with social media big data algorithms, to carry out research on mental illness early warning analysis technology.

This article mainly discusses these aspects. First of all, an overview of the theories related to social media big data. Then, research on psychological crisis early warning and related theories. In addition, the application of big data clustering algorithm in mental health analysis and early warning is also introduced. Finally, an experimental study was carried out around the mental health analysis and early warning system, and relevant experimental data and analysis conclusions were drawn.

2 Related Theoretical Overview and Research

2.1 Overview of Social Media Big Data Related Theories

The emergence of social media big data is accompanied by the development of network technology, from the traditional Internet era to the mobile Internet era, and this also provides a new opportunity for it. In this process, people can share and communicate after receiving information through various terminals such as mobile phones and computers. However, due to the openness and virtualization characteristics of the Internet, these communication platforms often have a large number of false news and rumors that affect user emotions, which in turn mislead the audience to misjudge and even cause social panic incidents. On the other hand, big data technology can effectively dig out massive amounts of data and analyze them, so as to draw corresponding conclusions.

In the era of big data, the emergence and development of social media has made the world more colorful. People receive information through a variety of channels such as the Internet and TV. In this case, a large amount of massive data has been generated. These huge, huge and valuable contents are gathered together to form a whole, which is the so-called "big database".

Social media refers to the use of digital technology to realize the exchange, sharing and delivery of information through the interaction between people and society. Social media refers to an online community established with individuals as the center. After users register, they obtain corresponding information and knowledge through information push and data analysis. With the rapid development of Internet technology and mobile communication technology and the rapid increase in the number of mobile Internet users, more and more people use mobile phones to browse news, etc. or spend a lot of time on social media platforms in fragmented time [4, 5].

People are gradually inseparable from social media in daily life, so social media big data analysis technology is indispensable in the construction and application of social information. Big data analysis uses a large number of valuable keywords to describe its internal laws and characteristics in a large amount of information. It mines and processes massive, unstructured, and large amounts of complex redundant data with low value density.

With the rapid advancement of information technology such as Internet technology and mobile devices, and the continuous expansion of the scope of human activities, people can obtain rich and diverse information resources through the platform of social media, and can meet their own needs in a short time.

Big data is proposed as an emerging concept and is developing rapidly. With the continuous advancement of network communication technology, smart phones have gradually become more powerful and upgraded, bringing convenient information services to people. Social media is represented by WeChat instant messaging software, Weibo, forums and other platforms are used frequently by users and grow rapidly, and they have become one of the most mainstream social media [6, 7].

The collection and analysis of social media big data is a dynamic development process. In this era of rapid change and strong interaction, it is necessary to constantly adjust and improve the massive amount of information collected in social networks. Therefore, it is particularly important to filter and process these large amounts of unstructured data.

2.2 Psychological Crisis Early Warning and Related Theoretical Research

In recent years, with the development of economy and society, a person's society, family, academic and working environment have become more complex and diverse, and the pressure of study, work, and life has made colleges and universities a high-risk group of mental disorders. The development and popularization of advanced technologies provide modern methods and means for colleges and universities to early warn students' psychological crisis ability, and also provide a theoretical and technical basis for the construction of mental health education and storage platforms.

Man is a unit of physiology, psychology and society. Man is not only a living body, but also a person with complex psychological activities and living in a certain social environment. Therefore, physical health and mental health are mutually integrated. When a person's physiology is sick, the psychology will inevitably be damaged, which will lead to depression, irritability and irritability, leading to long-term depression and even serious psychological disorders [8, 9].

Mental health is a positive and continuous attitude. In this state, a person can adapt well to the environment, his life is full of vitality, he can make full use of his physical and mental potential, which can be said to be mental health. Therefore, the level of human mental health can be roughly divided into three categories. The first category is normal general psychology, which is manifested as a good sense of humor, strong adaptability, good communication with others, the ability to carry out activities suitable for development, and ability to regulate emotions. The second is mild mental disorder, which is manifested as lack of happiness of peers, difficulty in getting along with others, poor self-care ability, and can return to normal after active adjustment or professional help. The third is a severe adjustment disorder. The patient's morbid psychological

manifestations are unable to maintain a normal life and work. If not treated in time, the condition may worsen and mental illness may appear.

Psychological crisis refers to the moment when a person faces sudden changes in the surrounding environment or major events in life, mental or behavioral obstacles, and shows confusion, leading to physical and mental imbalance. People’s own adaptability and mental balance are disrupted, not the stimulation itself.

If the psychological crisis cannot be contacted in time, it will develop into a serious psychological crisis, which may lead to suicide. Therefore, to deal with psychological crises, we must have a correct attitude, not be afraid, not evade, and actively guide and help patients to tide over all kinds of psychological crises [10, 11].

Psychological crisis warning is intended to predict and prevent psychological crisis. Psychological crisis intervention refers to a series of measures to help a person out of the current dangerous situation after discovering various psychological crises, and ultimately alleviate the crisis. In psychological crisis intervention, the principles of prevention, speed, and confidentiality should be strictly followed. The purpose of psychological crisis intervention is to help individuals regain their psychological balance. The psychological cognition model believes that the crisis originates from the misunderstanding of frustration and trauma. By changing the bad cognition, the individual can control the crisis, so the individual is guided to intervene in the psychological crisis. The psychological crisis early warning indicator system is shown in Fig. 1.

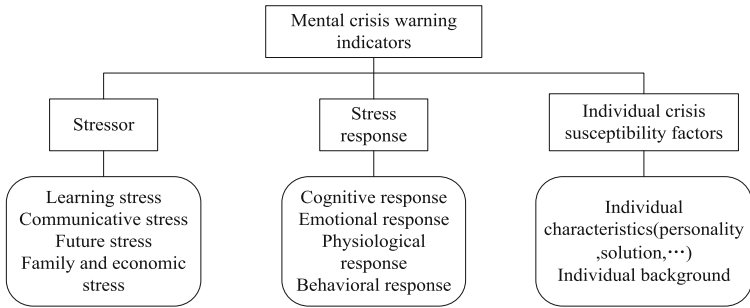


Fig. 1. Psychological crisis early warning indicator system

2.3 Application of Big Data Clustering Algorithm in Mental Health Analysis and Early Warning

The big data clustering algorithm needs to clarify the number of clustering categories in advance before performing cluster analysis on the data set. First, determine the number of categories to be subdivided, and select a representative point for each category, and perform the initial subdivision according to the principle of similarity within the category. Then try to change the position of the representative point to be optimized. Evaluate the value of the criterion function, repeat this process many times, and finally divide until the value of the criterion function is optimal. Through the application of big data clustering algorithm, to find the trend of each individual’s psychology, and find a certain law from

it, and finally complete the work of mental health analysis and early warning [12]. Part of the calculation formula is shown in (1), (2).

$$a_j = \frac{\sum_{u=1}^{m_j} c_u}{m_j} \quad (1)$$

$$F = \sum_{u=1}^j \sum_{c \in a_u} \|c - a_u\|^2 \quad (2)$$

Among them, c_u is the data point, j is the number of data points, a_j is the cluster center of the j th category, and m_j is the number of data points in the j th category. F is the criterion function. When its value reaches the minimum and remains stable, it means that the clustering work is completed.

3 Experiment and Research

3.1 Experimental Background

With the rapid development of economy and society, people's living standards continue to improve, but at the same time it also brings a lot of negative emotions. Among them, mental illness is one of the important reasons leading to the emergence or deepening of various mental crises. It causes damage to the individual's physical function or social function, and emotional reactions such as anxiety and depression, with a certain incubation period, and even serious consequences. When these problems are not resolved, it will affect social stability and the lives of the people. Therefore, timely monitoring and early warning of mental illness is very necessary.

3.2 Experimental Environment

In this experiment, the test system uses the B/S method to realize some related functions and applications of the psychological evaluation analysis and early warning system. The system also uses windows2015 64 database server, the back-end database uses relational database postgresSQL9.4, the operating system used by the WEB server is windows server 2015, and the server of the WEB publishing platform is ApacheHTTP. The client computer must be able to access the WWW server, and the browser must be IE6.5 or higher. In order to ensure the security of the system, a firewall needs to be deployed between the client and the server.

3.3 Experimental Process

In this experiment, combined with social media big data algorithms, a mental health analysis and early warning system was constructed. Through the analysis of the collected data, timely feedback of the individual's mental state in various time periods, to provide reliable data for monitoring the mental state, and prevent problems before they occur. Test the performance of the test system, including system reaction time, CPU utilization, and analysis results.

4 Analysis and Discussion

In this experiment, the performance of the test system was tested, including system reaction time, CPU utilization, and analysis result time. The test results are shown in Table 1.

Table 1. Test results of system performance

Concurrent number	Response time (s)	CPU utilization (%)	Analysis result time (s)
50	0.78	4.5	2
100	1.5	7.1	2.6
200	1.98	8.9	3.1
300	2.73	9.6	3.9
500	3.51	11.3	4.5

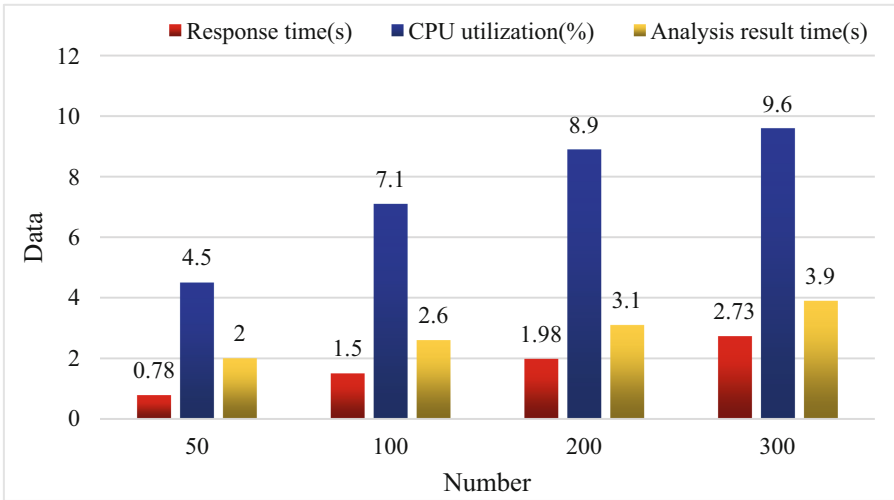


Fig. 2. Test results of system performance

It can be seen from Fig. 2 that when 50, 100, 200, 300, and 500 people visit the mental health analysis and early warning system together, the system response time is 0.78 s, 1.5 s, 1.98 s, 2.73 s, and 3.51 s, respectively. The analysis result time is 2 s, 2.6 s, 3.1 s, 3.9 s and 4.5 s respectively. It can be seen that when the number of concurrent users is less than 500, the response time of the system is shorter, and the time to get the results of mental health analysis is also shorter, so it can meet the basic needs of mental health analysis and early warning.

5 Conclusion

Facing complicated social life and work, people often have negative emotions such as anxiety and pessimism, and these negative emotions affect a person's healthy growth. And mental illness is also a topic that cannot be ignored. Therefore, how to monitor and warn mental illness in time has become a hot topic. Therefore, this article starts from a new perspective, combined with social media big data algorithms, to carry out research on mental illness early warning analysis technology, which has important practical significance and research value.

References

1. Juan, L., Luo, H.: The application of big data mining methods in the psychological early warning system of college students. *Chin. School Health* **039**(012), 1821–1824,1827 (2018)
2. Imran, M., Sial, M.H.: The role of stock exchange and public spending on economic growth in selected SAARC countries: a panel data analysis. *J. Crit. Rev.* **8**(1), 928–939 (2021)
3. Huang, X.: Relevant research on the mental health education platform for college students in the era of big data. *Sci. Popular* **000**(008), 46-47 (2018)
4. Ullah, H., Wan, W., Haidery, S.A., et al.: Spatiotemporal patterns of visitors in Urban Green parks by mining social media big data based upon WHO reports. *IEEE Access* **8**, 39197–39211 (2020)
5. Masih, J., Verbeke, W., Deutsch, J., et al.: Big data study for gluten-free foods in India and USA using online reviews and social media. *Agric. Sci.* **10**(3), 302–320 (2019)
6. Shu, K., Mahudeswaran, D., Wang, S., et al.: FakeNewsNet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data* **8**(3), 171–188 (2020)
7. Davis, J.L., Love, T.P.: Generalizing from social media data: a formal theory approach. *Inf. Commun. Soc.* **22**(5–6), 637–647 (2019)
8. Chan, J.L., Purohit, H.: Challenges to transforming unconventional social media data into actionable knowledge for public health systems during disasters. *Disaster Med. Public Health Prep.* **14**(3), 352–359 (2020)
9. Yoon, Y., Cho, M.: Social media bigdata analysis on Korean wave contents of Chinese. *J. Tourism Manage. Res.* **25**(2), 1–18 (2021)
10. Joshanloo, M.: Investigating the relationships between subjective well-being and psychological well-being over two decades. *Emotion* **19**(1), 183–187 (2019)
11. Hsieh, Y.P., Wei, H.S., Hwa, H.L., et al.: The Effects of peer victimization on children's internet addiction and psychological distress: the moderating roles of emotional and social intelligence. *J. Child Fam. Stud.* **28**(9), 2487–2498 (2019)
12. Pinto-Gouveia, J., Dinis, A., Gregório, S., Pinto, A.M.: Concurrent effects of different psychological processes in the prediction of depressive symptoms – the role of cognitive fusion. *Curr. Psychol.* **39**(2), 528–539 (2018). <https://doi.org/10.1007/s12144-017-9767-5>



Logistics Network Deployment Planning Under the Background of Big Data Technology

Meina Guo^(✉)

Dalian Vocational and Technical College, Dalian 116035, Liaoning, China
guomeina19820310@sina.com

Abstract. Logistics network points are nodes on the logistics network, which can be connected by transportation lines to form a logistics network. The distribution of logistics network nodes should be scientific, and measures should be adapted to local conditions, among which the overall deployment plan of the logistics network has the lowest cost. This article aims to discuss the logistics network deployment planning under the background of big data technology. First, it discusses the big data analysis system, and introduces the basis and principles of logistics node layout planning from two aspects: the theoretical basis of layout planning and the basic principles of regional logistics node layout planning. Secondly, it describes the details and models of the single-vehicle multi-path model, multi-time window and route template problem. The update process adds a combination of the global optimal solution and the optimal solution of this iteration. The results of the experiment show that the No. 47 car will be delivered at 8 am (set off work time), and the order delivery service will be completed before 18:00 (set off work time).

Keywords: Big data technology · Logistics network · Deployment planning · Urban logistics

1 Introduction

With the rapid development of urban logistics, urban logistics network system will also be formed. As an integral part of the logistics network system, the urban logistics network system is composed of nodes and channels [1]. Especially important is the internal structure of the urban logistics network, which is the foundation of the urban logistics network system. The location of the node determines the location, structure and design of the entire urban logistics network, which is the premise of network system planning [2]. The city network node is also a grand event for global logistics activities and service delivery services, as well as a large-scale new logistics provider. At the planning level, it is very important to first conduct a thorough investigation of the planning department [3, 4]. The design and construction of logistics network nodes in some cities is a weak link in the development of my country's logistics industry. Therefore, designing and constructing urban logistics node plates according to local conditions is the key to the development of modern logistics industry [5].

When planning multiple network nodes, the specific conditions of these networks and their surrounding areas must be fully considered to determine the most beneficial and optimistic logistics network points for the logistics company [6]. IQQ García uses the flexibility of UAS as a supplementary support for the medical logistics system under high pressure conditions by rapidly deploying an air transportation network. A logistics network model was defined and three scenarios were created based on the model and the current needs of Valencia (Spain). Flight tests were conducted in these scenarios, including urban areas and controlled airspace. The operation complies with the application requirements of the Specific Operational Risk Assessment (SORA) method recently adopted by the European Aviation Safety Agency (EASA). The flight was very successful, and medical supplies could be delivered quickly without any special infrastructure [7]. Li L proposed a new variant of the dual-objective capacity location routing problem (dual-objective CLRP), the flow capture location problem for fuel delivery (FCLP-FD), to simultaneously determine the location of hydrogen refueling stations (HFS) and routing hydrogen transport trucks decision. FCLP-FD aims to overcome the obstacles to deploying HFS. The HFS network was determined while maximizing the captured fuel demand flow and minimizing the total daily cost. A meta-heuristic algorithm for adaptive large neighborhood search (ALNS) was developed, and its performance in solving dual-objective optimization problems was tested [8]. The planning of logistics nodes plays an important role in achieving high efficiency and smooth business operations.

The research content of this article mainly includes the following aspects: introduced the basic structure of the urban logistics network, the layout of urban logistics network outlets, the concept and function of the urban logistics sector, and the management and operation of the logistics sector. Secondly, the Hadoop distributed computing system was launched, and finally the related research on logistics network outlet layout technology was carried out. Introduce the relative relationship between urban logistics and the process of network planning. And briefly explain the content and methods of logistics network research and analysis. At the same time, taking M City as an example, the establishment of urban logistics network network network point research, the technical planning basis of urban logistics network network point planning, combined with the logistics platform big data analysis system, recognizes the role of multiple time windows and route models in transportation planning.

2 Research on Logistics Network Deployment Planning Under the Background of Big Data Technology

2.1 Layout of Logistics Outlets

A logistics network is a junction of a logistics network and a connection point of multiple logistics lines in a logistics network. Logistics companies are the initiators and supporters of modern freight forwarding and the basis for the implementation of freight logistics [9, 10]. The construction of commercial logistics companies should be based on the development of major road intersections or expressways, and the development of logistics and wholesale products around expressways. This includes storage, processing, distribution, inspection, information management and technical services [11, 12].

With the rapid development of the logistics industry, the position of logistics outlets in the logistics network is becoming more and more important. Completing service quality and cost proposals, and understanding the location of local logistics outlets and the scale of local outlets are essential to achieve the efficiency of the logistics outlet layout system and ensure the smooth progress of the logistics system.

The operating scale of logistics outlets can be divided into two categories according to the number of bases. The first is the foundation of space. Only one logistics center has been opened in the entire area. This is the simplest case. The second is the first multipoint. This layout requires multiple logistics sites to be set up locally. This situation is more complicated, and the structural model and calculation process are also more complicated.

2.2 Big Data Analysis System

The big data analysis module is a distributed data analysis system based on Hadoop. The big data analysis module uses MapReduce to analyze available product tracking data, traffic conditions and local disaster information, and uses time series forecasting to provide real-time product demand forecasts for decision-making.

Use HDFS file sharing system to store a large amount of product search information. In terms of cost and performance, it is difficult to find the best big data tools. Using the HDFS extension, you will find that most data management cases have great flexibility and efficiency. On the other hand, users who have overcome statistical limitations with equipment do not have to worry too much about the code and how to use the code, so the same flexibility is the cause of inaccurate execution.

2.3 The Basis and Principles of Logistics Node Layout Planning

(1) Theoretical basis of layout planning

1) Modern logistics

When the company develops to a certain extent, the reduction of energy production capacity and the increase of productivity cannot bring profits to the industry, logistics has become a successful way for the company to reduce its capital capacity. The rapid growth of the logistics industry plays an important role in the development of the modern economy, and modern logistics capabilities have become one of the main competitors in this industry. The modern logistics industry has developed into a comprehensive service integration system integrating transfer, warehousing, loading and unloading, packaging, distribution, and information services to complete individual warehousing and transit logistics. The concept of logistics has also expanded from the distribution department to manufacturing, technical management and other departments. Modern logistics is developing multiple platforms and diversified technologies, emphasizing the integration of existing services, emphasizing logistics, network lines, system production connections, supply chain management and operations.

2) Space structure theory

Spatial structure theory is closely related to the research of location theory, because it is an important development of environmental science, but the focus of the current world structure process is slightly different from that of situational theory. The institutional

process is not the perfect state of any economic activity or event, but the interaction and social relations of many economic issues in this field. The field structure process is also called general science and world geography. Represents the size and continuity of objects and spatial phenomena, which may be reflected in the initial changes. Space science analyzes the general characteristics of the social system from different levels, and points out that the accumulation, fear and dependence of the system on the ground are the main factors that increase the world order and diversity. In the logistics node planning of the design area, the concept of logistics node location is very important for how to determine the scope of the logistics system and further reduce transportation costs.

(2) The basic principles of regional logistics node layout planning

1) The principle of market leadership

The current economic development of our country is dominated by the market system, and the work and management of enterprises must be market-oriented. Modern logistics is a work-based economic system, and its design and implementation must also be integrated across markets. Local node planning, transfer selection, project completion, project site selection, etc. all need to plan local development plans. It covers the prospects of major financial cities and markets across the country.

2) Location-oriented principle

The integration and connection points of different transportation systems form a comprehensive, strategic and state-of-the-art logistics node. The growth rate of the transfer determines whether the assembly can run smoothly. The structural design of the logistics department likes to choose a location that is easy to transfer, so there are many well-designed transfer methods that need to be integrated.

3) The principle of social orientation

The structure of local logistics departments must take into account system benefits and assume certain social responsibilities. It has the same financial orientation characteristics as the transfer material. In economically developed areas, in order to meet logistics needs and improve transportation efficiency, it is necessary to promote the economic development of developing countries and promote the development of economic integration. This is the goal of my country's national economy and logistics network construction.

3 Investigation and Research of Logistics Network Deployment Planning Under the Background of Big Data Technology

3.1 Current Status of Logistics Network Construction

The construction of logistics outlets is advancing rapidly. As of the end of 2020, M Province has built 25 freight stations, including 5 Class I ports, 5 Class II ports, 4 Class III ports, 6 Class II ports, and 5 freight stations. A number of logistics network projects are under construction and planning. The logistics infrastructure network is gradually taking shape. The daily cargo capacity of the port is 30,500 tons, and the average annual

cargo capacity of each cargo station is 5.65% of the local road freight volume. The throughput of the freight station is very small and cannot play the role of a distribution company, information center or management company.

Logistics information technology generally involves enterprises and management departments. Some companies develop and use logistics information systems with simple functions, such as train logistics management systems and Guanchebao logistics command systems. Some regional logistics platforms, such as Western Logistics Network, Zhongheng Logistics Network, and Quantum Logistics Network, have established public relations platforms in individual cities due to low automation and poor promotion, but they have not achieved the expected results.

3.2 Data Preprocessing

Suppose that x_{ij} represents the delivery vehicle from exit i to exit j , and C_{ij} represents the transportation cost from exit i to exit j . In this scheme, it is considered distance. This data can be calculated by the navigation distance in the map (to form a connection distance matrix). The optimization goal of the single-vehicle multi-path system is the shortest driving distance and the lowest transportation cost, as shown in Formula 1:

$$\text{Min}Z = \sum_i \sum_j c_{ij}x_{ij} \quad (1)$$

The opening or closing of multiple time windows is shown in formula 2:

$$TW = 0 \text{ or } 1 \quad (2)$$

The line template is turned on or off as shown in formula 3:

$$LT = 0 \text{ or } 1 \quad (3)$$

4 Investigation and Research Analysis of Logistics Network Deployment Planning Under the Background of Big Data Technology

4.1 Optimization Results of the Single-Vehicle Multi-path Model

The experimental results of the single-vehicle multipath model are shown in Table 1. The order of the test data is the order of the Order_50 file (provided by the company). There are a total of 27 tasks to be issued, the bicycle multi-path model, multiple time windows and route templates are not opened. At this point, it can be seen from the table that the order task in the Order_50 file is only completed by car #50. No. 50 will be shipped from 9:00 in the morning (set to be the first start), order service will be completed before 18:00 (set deadline), and the delivery plan will be executed. Order_50 calls the logistics route planning system according to the system order data, as shown in Fig. 1.

Table 1. Single-vehicle multi-path model optimization results

Vehicle ID	Serial number	Outlet ID	Order ID	Time
50	1	688852211403	10	8:08:41
50	2	688852211425	7	8:15:45
50	3	688852211429	3	8:29:03
50	4	688852211504	9	8:50:17
50	5	688852211709	21	9:20:48
50	6	688852211987	8	9:44:34
50	7	688852212015	4	9:59:41
50	8	688852212641	2	10:15:08
50	9	688852212666	5	10:40:10
50	10	688852212945	18	10:55:43
50
50	27	688852211356	22	17:04:03

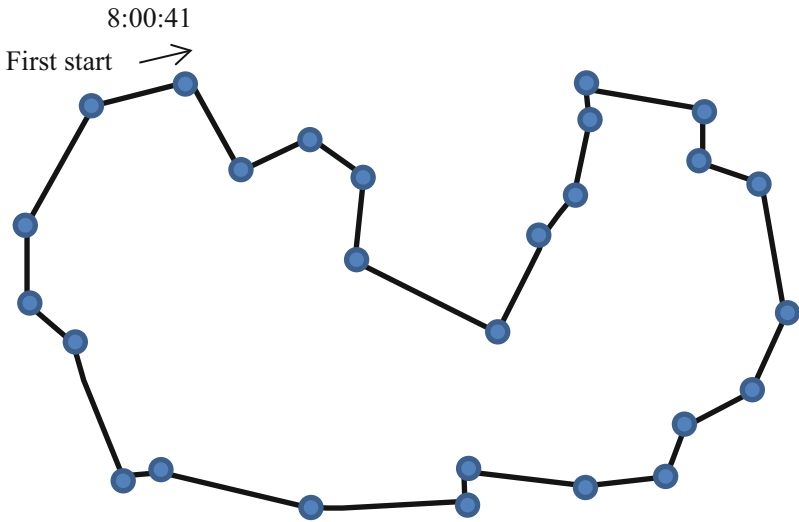


Fig. 1. Effect picture of single-vehicle multi-path experiment

4.2 The Influence of Multiple Time Windows and Route Templates on the Algorithm’s Shortest Path Optimization

Four modes of analysis and testing: time window not open and line template not open, time window open and line template not open, time window not open and line template open, time window open and line template open. Analyze the relationship between the shortest distances of order_10, order_20, and order_30 orders with multiple time windows and route templates, as shown in Table 2:

Table 2. The relationship between the shortest distance and multiple time windows and route templates

Experimental conditions	Order_10 shortest distance	Order_20 shortest distance	Order_30 shortest distance
Multiple time windows are not opened, line template is not enabled	18	25	50
Multiple time windows are enabled, and the line template is not enabled	25	36	78
Multi-time not open, line template is enabled	36	44	104
Open multiple time windows, enable line template	42	68	120

Under the principle of the same order quantity, if the time window and path model are not enabled, the path of the optimal solution obtained by the algorithm will be shorter; if the time window and path model are enabled at the same time, the path for the algorithm to find the optimal solution is the longest as shown in Fig. 2. Under the same test conditions, with the increase of the number of outlets, the longer the shortest distance, the influence of multiple time windows and line templates on the optimization algorithm can be seen.

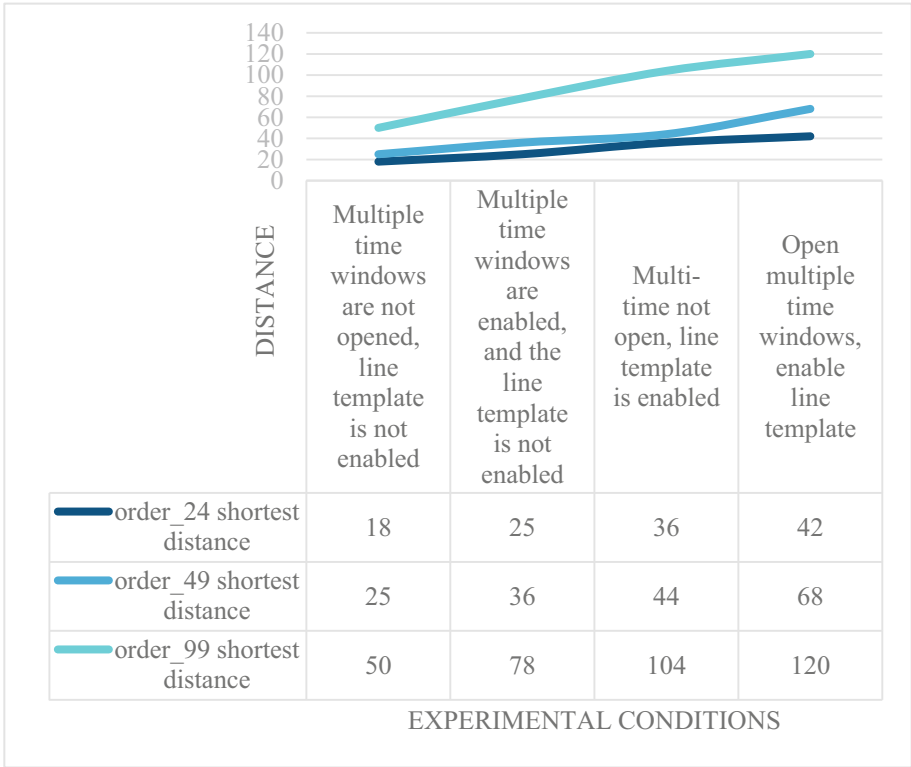


Fig. 2. The relationship between the shortest distance and multiple time windows and route templates

5 Conclusions

With the rapid development of the global system, logistics plays an increasingly important role in the social system. The cost of logistics services is an important factor that affects the development of investment fields and systems, and directly determines the difference in revenue costs between countries and regions and the competitiveness of enterprises. Modern logistics is a key strategy to support sustainable economic growth, increase investment areas, improve social and economic benefits, reduce social costs, and maximize the use of social resources. This article combines the existing scientific research and the main issues of logistics development, from logistics network to logistics network planning, transportation planning, transportation organization, planning process and other aspects to examine the components of logistics network planning. It is necessary to improve and correct the differences in understanding of the basic assumptions of the urban strategic planning system, and formulate strategic plans for logistics network planning and direction suitable for current and future economic conditions and markets. The results of this study provide a basis and platform for further research on the proposed urban logistics network planning plan, and provide a reference for the planning of urban logistics network nodes to put forward more thoughtful suggestions.

References

1. Hofman, W.J.: A methodological approach for development and deployment of data sharing in complex organizational supply and logistics networks with blockchain technology. *IFAC-PapersOnLine* **52**(3), 55–60 (2019)
2. Chen, H., Jonchay, T., Hou, L., et al.: Multifidelity space mission planning and infrastructure design framework for space resource logistics. *J. Spacecr. Rocket.* **58**(2), 1–14 (2021)
3. Soshino, Y., Kuroda, A., Miyata, A.: Review of the domestic utilization of pre-positioned international medical relief equipment in Japan. *J. Nat. Disaster Sci.* **38**(2), 165–178 (2017)
4. Moshood, T.D., Nawanir, G., et al.: Digital twins driven supply chain visibility within logistics a new paradigm for future logistics. *Appl. Syst. Innov.* **4**(2), 29 (2021)
5. Yu, W., Lin, H., Wang, Y., et al.: A ferrobotic system for automated microfluidic logistics. *Sci. Robot.* **5**(39), eaba4411 (2020)
6. Liang, W., Zheng, M., Zhang, J., et al.: WIA-FA and its applications to digital factory: a wireless network solution for factory automation. *Proc. IEEE* **107**(6), 1053–1073 (2019)
7. García, I.Q.Q., Vélez, N.V.V., Martínez, P.A.A., et al.: A quickly deployed and UAS-based logistics network for delivery of critical medical goods during healthcare system stress periods: a real use case in Valencia (Spain). *Drones* **5**(1), 13 (2021)
8. Li, L., Chami, Z.A., Manier, H., et al.: Incorporating fuel delivery in network design for hydrogen fueling stations: formulation and two metaheuristic approaches. *Transp. Res. Part E Logistics Transp. Rev.* **152**(1), 102384 (2021)
9. Bonebrake, V.: Designing the ‘Perfect’ logistics network. *Supply Chain Brain* **21**(1), 72 (2017)
10. Skender, H.P., Zaninovi, P.A., Loli, A.: The importance of logistics distribution centers as nodes in logistics networks. *Pomorstvo* **33**(2), 149–157 (2019)
11. Lam, C.Y.: Optimizing logistics routings in a network perspective of supply and demand nodes. *CEJOR* **29**(1), 357–377 (2019). <https://doi.org/10.1007/s10100-019-00653-w>
12. Rama, V.K.: Logistics sector way ahead - customer value creation. *Indian Railways* **63**(1), 188–189 (2019)



Power Grid Parameter Acquisition System Based on the Fusion of Intelligent Unit and Edge Computing

Dawei Wang¹, Jianjun Liu²(✉), Bo Zhao⁴, Xiaofeng Xu³, Kenan Hou¹, Bing Ma¹,
Wei Tang⁴, Changyuan Bao⁴, Shuwei Hou¹, and Chunyang Dong¹

- ¹ Beijing Smart-Chip Semiconductor Technology Co., Ltd., China Sciences Cloud Valley Park,
No. 79 Shuangying West, Changping District, Beijing 102200, China
- ² State Grid Key Laboratory of Power Industrial Chip Design and Analysis Technology,
Beijing Smart-Chip Microelectronics Technology Co., Ltd., C-3, Dongsheng Science and
Technology Park, No. 66, Xixiaokou Rd, Haidian District, Beijing 100192, China
- ³ State Grid Jilin Electric Power Company Limited, No. 10388, Renmin Street,
Nanguan District, Changchun 130000, China
- ⁴ State Grid Jilin Electric Power Company Limited, Jilin Power Supply Company, No. 124,
Jilin Street, Chuanying District, Jilin 132000, China

Abstract. At present, with the rapid development of intelligent technology, people have higher and higher requirements for power quality. Edge computing, as an important tool for effectively processing power system data and simulating related processes, realizing machine learning functions, and performing mathematical modeling, is widely used in various fields. The integration of intelligent units and edge computing is a strong technology with a wide range of applications, involving multiple disciplines such as power systems, network communications, and computers. At the same time, in today's power system, in some cases, grid parameters and power consumption information cannot be monitored and read in real time, which seriously affects the operational safety of the entire grid to a large extent. Therefore, in order to better guarantee the quality of power supply and reduce energy consumption, research on grid parameter collection systems based on the integration of smart units and edge computing is carried out. This paper adopts experimental analysis method and data analysis method, which is intended to improve the stability and controllability of the grid parameter collection system by combining the advantages of intelligent units and edge algorithms. According to the experimental results, it can be seen from the experimental data that the output value of the power acquisition circuit is slightly higher than the calculated value, indicating that there is a measurement error, but the error is relatively small. In addition, if the power consumption of the test platform itself is removed, the two data sets will be closer, and it can be concluded that the power data acquisition and processing circuit has higher measurement accuracy.

Keywords: Intelligent unit · Edge computing · Parameter acquisition · Data processing

1 Introduction

Nowadays, with the continuous improvement of the level of intelligence, people have put forward higher and higher requirements for power quality and reliability. The most critical one is based on edge computing technology. The research on the fusion algorithm of intelligent unit and edge computing is a new subject, which has made great progress in theory and application technology. At the same time, traditional grid parameter collection methods can no longer meet the power system control requirements with high power quality and stability requirements. Therefore, by combining the advantages of smart units and edge algorithms, the stability and controllability of the grid parameter acquisition system are improved.

At present, research results on edge computing and distribution network data acquisition systems are very rich. For example, Hong Hui believes that it is necessary to work hard to overcome the problem of artificial intelligence unit teaching and optimize the teaching process, which is conducive to cultivating students' thinking ability [1]. Qi Yanli pointed out that in the future, 5G mobile communication networks will usher in the challenges of skyrocketing mobile traffic and new services with high backhaul bandwidth and low latency. Mobile edge computing MEC is expected to solve the above problems [2]. He Zhenting proposed that the operation principle of power system monitoring and data acquisition is closely related to the functions of each part of the two important systems of monitoring and data acquisition [3]. Therefore, the research on the grid parameter collection system under the integration of smart units and edge computing is in line with the development trend of the times, and has important significance and research value of the times.

This article mainly discusses these aspects. First of all, relevant explanations are made on the meaning of smart units and edge computing. Then, the collection of power grid parameters and its related content are explained. Finally, the experimental research is carried out for the grid parameter acquisition system under the integration of intelligent unit and edge computing, and relevant experimental results and analysis conclusions are drawn.

2 Research on Grid Parameter Acquisition System Based on the Integration of Intelligent Unit and Edge Computing

2.1 Intelligent Unit and Edge Computing

The intelligent unit can provide unique services according to the location of the place and the characteristics of the facility, and the service agent is located in the unit that provides basic services in the intelligent room. In this way, the level of service can be clarified. Some basic services can be provided separately and equally from a large platform. Each entity must provide unique services based on its own facilities and functions. In addition, there is such an explanation. The smart unit is the work or living space in the smart space, with basic service agents and unique facilities, which can provide unique services [4, 5]. The intelligent unit-smart space model diagram is shown in Fig. 1.

The smart unit is an integral part of the smart space. The smart unit communicates with the space through an agent, and the agent in the unit requests basic services from

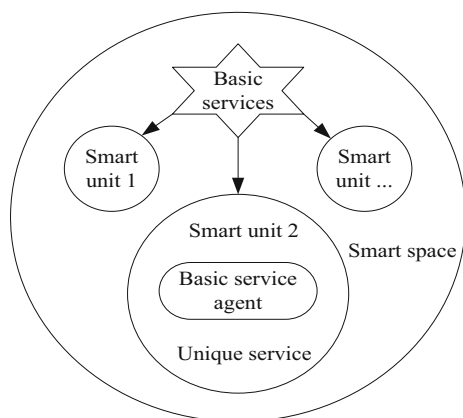


Fig. 1. Smart unit-smart space model diagram

the smart space. At the same time, the scope of the intelligent unit determines the limits of the equipment managed by the unit management center, and also the limits of the permissions granted to users.

The software layer model of the intelligent unit is divided into a perception layer, a network layer, a data processing layer and a service layer. The perception layer includes the sensing equipment of the intelligent unit, and the physical equipment of the intelligent equipment is divided into two parts: agent equipment and stand-alone equipment. The unique equipment is mainly testing equipment. A single device is the source of the basic data information of the intelligent unit. The data collection technology is mature and has a wide variety, including: radio frequency identification, infrared beacon, two-dimensional code, etc. Sensors or other sensor devices record large amounts of environmental information in real time, collect data, and transmit wireless or wired data. Communication technology transmits short distances to central nodes or network equipment. In principle, the area covered by the sensor array is not limited by the area, and the assembly can be completed quickly, which has great advantages in assembly time, convenience and reliability [6, 7].

The network layer mainly implements the following two functions: One is to discover the network connection, that is, the underlying networking. The second is heterogeneous network conversion, that is, the conversion between the lower-layer network and the upper-layer network, and the data conversion problem between the lower-layer sensor network must be solved at the same time; the function of the network layer mainly depends on the realization of the intelligent gateway and the overall framework of the intelligent gateway.

The data processing layer is the management center of the intelligent unit, and the equipment is not restricted. According to the specific conditions of the intelligent unit, gateways, routers, etc. can all be used as management centers to implement data preprocessing, proxy, and data storage functions. The network layer and the data processing layer are the two layers of the software structure, which can be the same device in actual situations.

At present, advanced data processing technologies centered on edge computing are being developed and widely used. Edge computing is defined as an open, distributed platform that integrates basic network, computing, storage, and application functions at the edge of the network close to the source of things or data. In fact, edge computing is a new green model. By converging the five resources of network, computing, storage, application, and intelligence at the edge of the network, the performance of network services is improved, network control is opened up, and new business forms similar to the mobile Internet are stimulated.

Edge computing is a new type of distributed computing method that can realize stable and fast data processing and real-time execution control at the edge of the network. Introduce advanced computing technology into the active power distribution network to improve real-time IT performance to meet CPS's requirements for low-latency services [8, 9].

2.2 Grid Parameter Collection

Now that we have entered the information age, big data, the Internet, and cloud computing have all achieved rapid development. Real-time monitoring and automatic control of power systems also reflect the requirements of the development of the times. However, the traditional manual measurement method has many drawbacks: low efficiency and large error, slow acquisition speed and low accuracy. With the continuous integration of modern information technology and communication networks and the rapid popularization and application of computer technology in various industries, under the unstoppable trend, the development of grid parameter collection has unlimited prospects.

The power system is a complex and huge spatial network composed of various generators, transformers, and high-voltage transmission lines. Power grid parameter collection is to use sensors and microprocessor technology to monitor the power grid operating characteristics in real time. After connecting to the network platform, the data of the power system data center can be collected in real time and appropriate judgments can be made to monitor and control the collection of emergency events in the power system [10, 11].

Grid parameters refer to data such as voltage and current during the operation of the power system, which reflect the loss of electrical energy. Manual collection, first manually collect field data, and then send the recorded data to the corresponding experts for processing and analysis. Finally, the responsible experts will put forward their opinions, and the registered personnel will log on to the website for debugging. This method has many subjective factors and cannot process data in time, and because of its many shortcomings, it has not been widely used.

The basic task of power measurement is to accurately and quickly collect and transmit power grid parameters, and through the data acquisition system, real-time monitoring of various types of voltage and current signals or power meter output waveforms.

The grid parameter system does not need to manually input data on site, and can directly send the data to the data center, and then directly processed by relevant experts and sent back. Its advantage is that it is not subject to external interference, so it has stable characteristics. Although there will be some delay when the distance is far, but with the development of technology, the delay time gradually becomes negligible, and

remote monitoring and control can be realized. But the disadvantage is that the initial installation cost is very high, and it is more difficult to troubleshoot problems when the wired network is damaged [12].

These collection points are connected through a wireless network link, and finally wireless communication between them is established. The remote sensing of the wireless network is inseparable from the GPRS technology, and its real-time performance is strong, which can ensure the safety of users. It also has the advantages of wide coverage, convenient access and connection. Compared with wired networks, the disadvantages are poor stability, error-proneness, and poor interference suppression capabilities.

3 Experiment and Research

3.1 Experimental Environment

In this experiment, to ensure reliable transmission, a stream socket is used to implement TCP communication. The host computer software is developed using Microsoft Visual C++ 5.0. Winsock is a widely used network programming interface on Windows and supports multiple protocols. It is widely used in TCP/IP programming. The data socket is based on the UDP protocol, and the data stream can be transmitted in both directions, but the reliability and stability are slightly worse, and data can be transmitted even if there is no connection. The original socket interface has low requirements on the protocol layer, and can allow access through IP, ICMP, etc., so it is widely used to review new protocols to access new devices on existing servers. The electric quantity sensor circuit adopts ADE7756 of American ADI Company. Due to the use of digital circuits to process and transform the collected signals, the time and temperature stability of the entire circuit is very high. The specific formula expression is shown in (1) (2).

$$v(r) = V_0 + \sqrt{2} \sum_{j=0}^{\infty} V_j \sin(j\theta r + \chi_j) \quad (1)$$

$$o(r) = O_0 + \sqrt{2} \sum_{j=0}^{\infty} O_j \sin(j\theta r + \gamma_j) \quad (2)$$

Among them, $v(r)$ is the instantaneous voltage, V_0 is the average voltage, V_j is the effective value of the j th harmonic of the voltage, and χ_j the effective value of the j th harmonic of the current. $o(r)$ is the instantaneous current, O_0 is the average current, and γ_j is the j -th harmonic phase angle of the current.

3.2 Experimental Process

In this experiment, the test platform for the electrical data acquisition and processing circuit includes AC voltage stabilization devices, ammeters, voltmeters, and power factor measuring instruments. The test platform verifies the performance of the power data acquisition and processing circuit. Use 110 V, 150 V, 180 V, and 220 V AC voltages in the AC voltage stabilization device, and use measuring instruments such as ammeters

to detect the current and voltage of the circuit, and calculate the energy data per unit time (kW·h) according to the measurement time. Compare this calculation result with the output value of the electric quantity data acquisition and processing circuit to verify its accuracy. Some experimental results are shown below.

4 Analysis and Discussion

In this experiment, AC voltages of 110 V, 150 V, 180 V and 220 V were used in the AC voltage stabilization device to detect the current and voltage of the circuit, and calculate the energy data per unit time (kW·h) based on the measurement time. A calculation result is compared with the output value of the power data acquisition and processing circuit. The test results are shown in Table 1.

Table 1. Test results of grid parameter acquisition system

Voltage (V)	Electric current (A)	Measure time (h)	Calculated value of electric energy (kW-h)	Electric energy output value (kW-h)
110	0.31	0.4	0.198	0.241
150	0.39	0.8	0.271	0.397
180	0.43	1.2	0.523	0.596
220	0.51	1.6	0.68	0.712

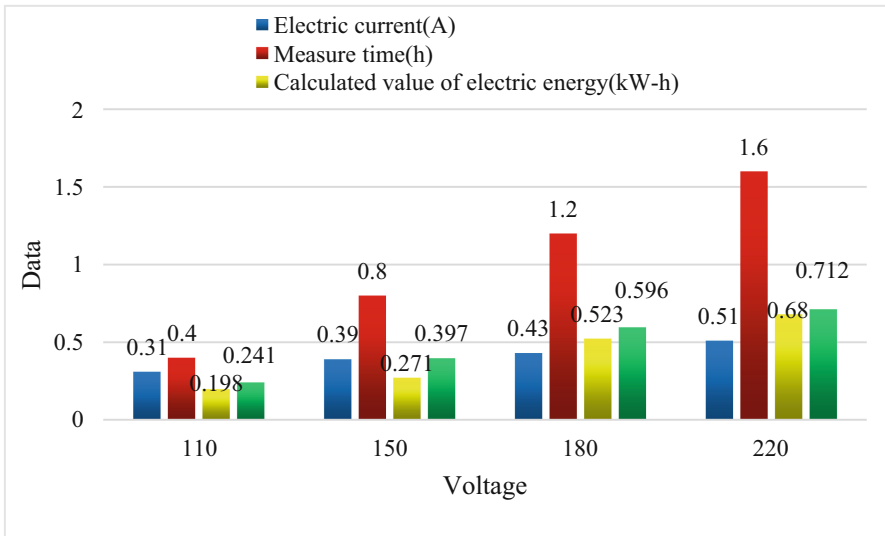


Fig. 2. Test results of grid parameter acquisition system

It can be seen from Fig. 2 that when the AC voltage stabilization device uses 110 V, 150 V, 180 V, and 220 V AC voltages, the output values of the power collection circuit are 0.241, 0.397, 0.596, and 0.712, respectively, and the calculated values of the power collection circuit are 0.198, 0.271, 0.523 and 0.68. It can be seen from these data that the output value of the power acquisition circuit is slightly higher than the calculated value, indicating that there is a measurement error, but the error is relatively small. In addition, if the power consumption of the test platform itself is removed, the two data sets will be closer, and it can be concluded that the power data acquisition and processing circuit has higher measurement accuracy.

5 Conclusion

With the development of smart grid and distributed power generation technology and the continuous expansion of application fields, the technology and requirements for grid parameter collection are getting higher and higher. In today's power system, edge computing is an important method that can accurately and reliably calculate the transmission network. At the same time, the fusion algorithm of smart units and edge computing is a current research hotspot. After the mathematical relationship between the two is established, how to effectively integrate the two and use their advantages to collect grid parameters more efficiently has become a hot topic and the focus of the majority of scholars. Therefore, this paper integrates intelligent units and edge computing related technologies to conduct research on the grid parameter acquisition system.

Acknowledgment. This work is supported by the Science and Technology Research Project of State Grid Corporation of China, project No. 546816190007: Research on the key technology of microcontroller for intelligent acquisition unit based on edge computing and Near-Subthreshold.

References

1. Hui, H.: Teaching strategy and practice analysis of the artificial intelligence unit of information technology in junior high schools (2020–47), pp. 29–30 (2021)
2. Yanli, Q., Yiqing, Z., Ling, L., et al.: Future 5G mobile communication network integrating mobile edge computing. *Comput. Res. Dev.* **55**(3), 478–486 (2018)
3. Zhenting, H.: Research on power system monitoring and data acquisition system. *China Strateg. Emerg. Ind.* **08**, 185–186 (2017)
4. Thejaswini, N., Lahari, D., Baba, S., et al.: Monitoring and automated hydroponic system using smart power unit. *Int. J. Control Autom.* **13**(4), 47–54 (2020)
5. Sun, Z., Xing, X., Tian, W., et al.: An optimized clustering communication protocol based on intelligent computing in information-centric internet of things. *IEEE Access* **7**(99), 28238–28249 (2019)
6. Sabir, Z., Raja, M., Arbi, A., et al.: Neuro-swarms intelligent computing using Gudermannian kernel for solving a class of second order Lane-Emden singular nonlinear model. *AIMS Math.* **6**(3), 2468–2485 (2020)
7. Muhammad, Y., Khan, R., Raja, M., et al.: Design of fractional swarm intelligent computing with entropy evolution for optimal power flow problems. *IEEE Access*, 1 (2020)

8. Zs, A., Mazrb, C., Ymcd, E., et al.: Design of neural network based intelligent computing for neumerical treatment of unsteady 3D flow of Eyring-Powell magneto-nanofluidic model. *J. Market. Res.* **9**(6), 14372–14387 (2020)
9. Xu, G., Yang, H.-Y., et al.: Design and implementation of a method for temperature parameter collection inside acupoint. *Zhongguo zhen jiu = Chinese acupuncture moxibustion* **39**(8), 901–906 (2019)
10. Ke, R., Li, Z., et al.: Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow. *IEEE Trans. Intell. Transp. Syst.* **20**(1), 54–64 (2019)
11. Xiao, G., et al.: Sparse representation of tropospheric grid data using compressed sensing. *GPS Solutions* **25**(3), 1–15 (2021). <https://doi.org/10.1007/s10291-021-01120-3>
12. Razavi-Far, R., Farajzadeh-Zanjani, M., Saif, M., et al.: Correlation clustering imputation for diagnosing attacks and faults with missing power grid data. *IEEE Trans. Smart Grid* **11**(2), 1453–1464 (2020)



Intelligent Analysis and Processing System of Financial Big Data Based on Neural Network Algorithm

Wei Shi^(✉)

Wuhan Railway Vocational College of Technology, Wuhan 430205, Hubei, China
Shiwei202111@126.com

Abstract. With the continuous development of computer technology, data mining has gradually become an indispensable part of enterprise information management. Traditional enterprises have been unable to meet the needs of modern economy and Society for talents, and big data, as a new technology, appears in the field of internal information processing. Financial analysis is to comprehensively evaluate the company's daily business activities, comprehensively grasp and reasonably allocate resources, so as to improve efficiency and reduce costs. Therefore, using neural network algorithm to realize big data prediction is more and more important and practical for modern enterprises. This paper first expounds the concept of financial big data and the characteristics of data intelligent analysis, and then introduces the application of neural network model. Then, it will study based on financial big data, design a complete and effective neural network intelligent algorithm system that can be applied in practice, and test the effectiveness of financial big data intelligent analysis and processing system. Finally, the experiment shows that the difference between the predicted value and the actual error value of the system data is small, and the effectiveness is very high. At the same time, the performance of data processing meets the needs of users, which can improve work efficiency and reduce labor cost.

Keywords: Neural network algorithm · Financial big data · Intelligent analysis · Analysis and processing

1 Introduction

With the development of computer technology, there are higher and more complex requirements for information processing in the era of big data. Artificial intelligence plays an increasingly important role in enterprises. The traditional manual management mode can no longer meet the needs of high efficiency, high quality and high-performance computing power in the modern economic era [1, 2]. Therefore, more advanced, efficient and intelligent systems are needed to improve work quality and reduce costs. At the same time, with the continuous expansion of network technology and computer application fields, enterprises need more and more data storage and use methods in the process of

financial analysis, and financial analysis is an indispensable part of data mining and prediction. People pay more and more attention to financial big data [3, 4].

Many researchers have investigated and applied the role of neural network algorithm in the field of finance. For the processing of accounting big data, there are some research results at home and abroad, and the research on it started earlier abroad. In the 1990s, General Electric Company first combined computer technology with traditional manual bookkeeping to carry out budget management, cost control and salary calculation. Subsequently, Toyota Motor manufacturing company of Japan also began to prepare annual statements manually and implement the method of budget implementation to realize internal audit [5, 6]. At the end of last century, the financial analysis circles all over the world have begun to make a breakthrough in using artificial intelligence to process accounting data. China also has a certain degree of reference for accounting information processing. China held the “big data research conference” in Beijing and established relevant academic organizations and institutions to promote the development and achievement transformation of new technologies such as big database and cloud computing [7, 8]. The above research has laid the foundation for this paper.

This paper will introduce the related contents involved in the intelligent financial system based on neural network algorithm from three aspects: the concept of financial big data, intelligent analysis and decision support. Firstly, this paper expounds the role of accounting information system in enterprise management, then analyzes and designs its internal control mechanism, and then puts forward corresponding measures for data processing process. Finally, based on intelligent algorithm, a complete and effective financial analysis system with good performance and high prediction accuracy is proposed.

2 Discussion on Intelligent Analysis and Processing of Financial Big Data Based on Neural Network Algorithm

2.1 Overview of Financial Big Data

The basic function of financial accounting is to record and report the economic movements that have occurred or will form in the past, and to provide various relevant information, including cash flow, balance sheet, total profit and so on, for all stakeholders. With the gradual deepening of the application of computer technology in modern society and the advent of the big data era, people have a deeper and deeper understanding of it. In the traditional sense, it mainly takes a short time from collection, processing and analysis to decision analysis, and it is easy to realize automatic processing. In the business process, enterprises will produce a large number of financial data, which can often reflect the future development of the company. For financial accounting, the analysis and processing of various financial information can help us better predict and understand the market competition situation of the industry [9, 10]. Financial management is the core of an enterprise. Financial data analysis is to integrate and process these information through computer technology and with the help of certain algorithms. For big data, it mainly has the following characteristics: first, it has diversity. Second, high value density. Third, strong timeliness and fast renewal. Fourth, it is unstructured and unstable.

Fifth, it is multidimensional and has little non relevance. Sixth, it is easy to expand and closely connected with cross departmental cooperation [11, 12].

2.2 Intelligent Analysis of Financial Data

Financial analysis is based on the enterprise's financial and accounting information and takes the enterprise's operation and financial activities as the object to analyze the enterprise's operating results, financial status, cash flow and other financial data, so as to help the enterprise managers make correct decision-making and selection, capital operation and financing planning, and help other information users make correct evaluation of the enterprise predictive judgment. Financial analysis provides necessary decision-making information for information users. Based on the results of financial analysis, we can make decisions scientifically and according to the basis, avoid blind decisions, and improve the possibility of making correct decisions. It is the basis of financial decisions. Through financial analysis and prediction based on the analysis results, we can grasp the possible dynamics of future development, have a deeper understanding of uncertainty, foresee the possible risks after the implementation of various decisions, promote the making of optimal decisions and help the smooth realization of objectives. Generally speaking, the basis of financial decision-making is to adopt appropriate methods to analyze historical financial data, so as to obtain and abstract the understanding and assumptions of future development trend. Any continuous operation enterprise participates in the relevant market because of operating a certain business. As a market participant, most of the markets faced by the enterprise are full of competition. Even if it occupies a "blue ocean" that needs to be further developed, it needs to establish a competitive advantage in time to prevent new entrants from participating in the competition. Enterprises are closely related to the survival and development of enterprises from fund-raising, long-term asset investment, working capital investment and income distribution from business activities. Therefore, through financial analysis, the observation and control of the enterprise's capital and financial situation can enable the enterprise managers to have an accurate understanding of the enterprise's financial situation, operating results and cash flow, and make judgments and decisions on financing, investment, development and contraction in time.

2.3 Neural Network Technology

2.3.1 Emergence and Development of Neural Network

Neural network is a new concept. It began to develop and spread rapidly in the 1950s. In 1975, the Massachusetts Institute of technology was the first to put forward the idea of combining manual and computer to process information. With the rapid growth of large-scale artificial simulation of human brain and the improvement of human brain cognitive ability, people have a more profound and comprehensive understanding of artificial intelligence. Neural network is processed by a large number of neurons through simulation and artificial methods. It has developed and formed a new branch based on imitating the characteristics of human brain thinking and language understanding. It consists of a large number of neurons interconnected to build a technology with

the characteristics of self-learning function, parallel processing ability and strong fault tolerance, and can automatically adapt and control the simulated human brain nervous system.

2.3.2 Algorithm

The final output result of artificial neural network can not be infinite. It is generally limited to a certain range, so there needs to be a function to limit the output - activation function. It can also be called a suppression function. This name is used because its main function is to suppress the input. Its output has a certain range, and completing the conversion from input to output is called an activation function. The output size of artificial neuron is limited, and its range is usually limited to the interval from 0 to 1, or the interval from -1 to 1. The above description can be explained by mathematical expressions.

$$\beta_k = \sum_{i=1}^n w_{ki}x_i \quad (1)$$

$$v_k = net_K = \beta_k - \theta_k \quad (2)$$

$$y_k = f(v_k) \quad (3)$$

The input vector is x_1, x_2, \dots, x_n , the P weights of neuron K are represented by $w_{k1}, w_{k2}, \dots, w_{kn}$, the product of input vector and weight is represented by β_k , and the sum of neuron K weighted input is represented by Y_k .

3 Experiment

3.1 Structure Module of Financial Big Data Intelligent Analysis System

Figure 1 is the module structure diagram of financial big data intelligent analysis system. Neural network algorithm has been widely used in the field of financial analysis, but because artificial neural network has the characteristics of nonlinearity and self-organization, it is necessary to establish a certain model to solve these problems when classifying data sets. The overall functions of the system include data entry, selection of artificial neural network algorithm and calculation of training parameters. In the input pattern recognition stage, artificial intelligence technology is used to preprocess the historical economic activities of enterprises. In the output layer, neurons convert the original information into standard form and store it in the big data center for processing, and automatically generate structured language files of corresponding types and formats according to the characteristics of different types of data.

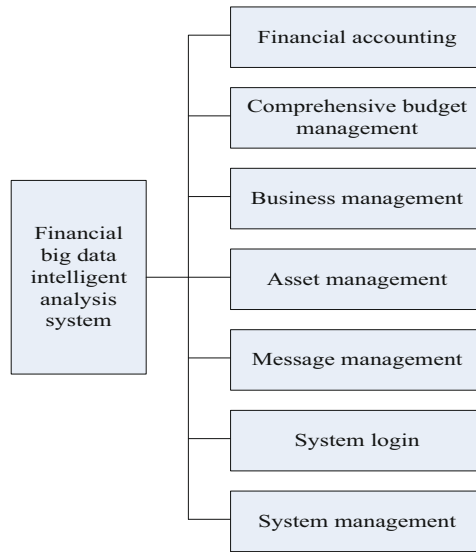


Fig. 1. Financial big data management system module

In the system login module, when an enterprise needs to obtain financial data, it first analyzes it through manual input and stores the information. Then, the main window function is used to store the output value and required time of each layer in the database. Set an index library in the main window to query whether each attribute variable belongs to the corresponding level subdirectory in the table and calculate the weight. If it does not exist, it will directly call relevant tables to find, compare, select and modify data types. If necessary, you can also sort them according to different classification functions.

3.2 Financial Data Collection

In the process of data collection, we must first complete the preprocessing of the original financial information and convert it into accounting books and statements that can be recognized by the computer. For enterprises, the collected accounting big data is very important, but the processing of these data is very complicated and complex. The data acquisition part is the core of the whole system, which is mainly completed manually, including preprocessing, feature extraction and standardization of the original data. First, input the original voucher and generate the corresponding report, then transfer all aspects of the report to the server through network transmission technology to meet the needs of different users, and finally get the results after the client sends the report to the system.

4 Discussion

4.1 System Effectiveness Analysis Based on Neural Network Algorithm

Table 1 shows the data results of the effectiveness test of the system.

Table 1. The expected and the actual values of testing

Number of tests	Expected value	Actual value	Predicting error values	Predicting error rate values
1	4.2563	4.5896	0.3306	0.077
2	3.3658	4.2547	0.8889	0.264
3	4.3658	4.1578	-0.208	-0.047
4	2.4879	3.5747	1.0868	0.436
5	3.2147	3.6987	0.484	0.150

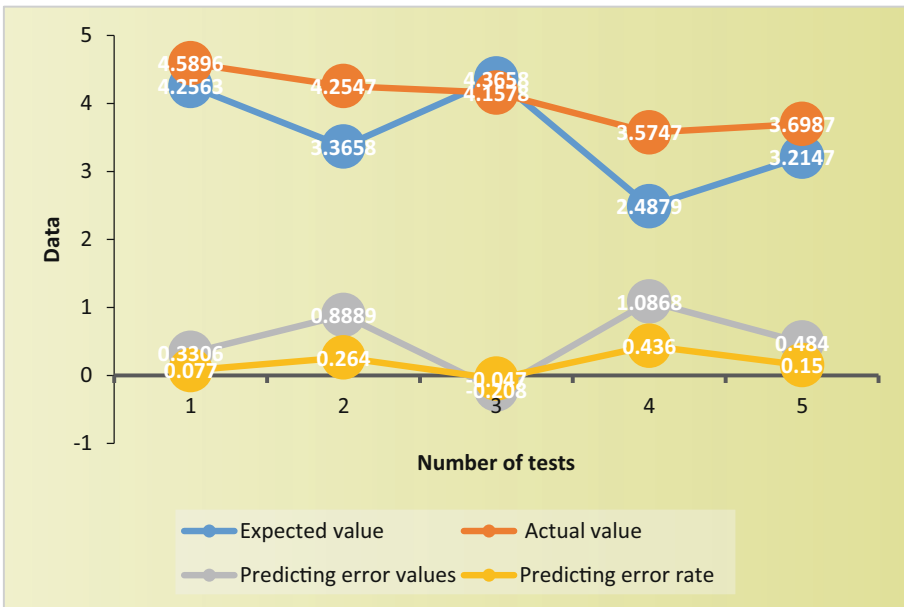


Fig. 2. System effectiveness test

System performance test is mainly to test the accuracy and reliability of data. Through data analysis, we can find out what problems exist in the enterprise production process and what impact may occur in these aspects. In this paper, the output of the system is compared with the actual detection value through the simulation function, and it is found that the error between the two is within the required range. As can be seen from

Fig. 2, the difference between the predicted value and the actual error value of the system data is small. It can be seen that the effectiveness of the system is very high, and the performance of data processing meets the needs of users.

5 Conclusion

This paper uses artificial neural network algorithm for financial prediction. Through the analysis of the actual situation of enterprises, it is found that many big data contain a lot of useful value information. Firstly, starting with the characteristics of big data, the traditional accounting work mode is changed to artificial intelligence, computer technology and the Internet, and then a variety of research methods such as mathematical modeling method and grey correlation method are used in model design to form a complete, effective and reasonable financial prediction process based on neural network algorithm and other relevant theoretical knowledge. By analyzing the internal structure of the enterprise, it is found that the system has strong real-time performance.

References

1. Zhou, C.: Algorithm design of early warning seatbelt intelligent adjustment system based on neural network and big data analysis. *Math. Probl. Eng.* **2020**, 1–7 (2020)
2. Zhang, C., Liu, X.: Feature extraction of ancient Chinese characters based on deep convolution neural network and big data analysis. *Comput. Intell. Neurosci.* **2021**(3), 1–10 (2021)
3. Cerquitelli, T., Quercia, D., Pasquale, F.: Visualizations of deep neural networks in computer vision: a survey. In: [Studies in Big Data] *Transparent Data Mining for Big and Small Data*, vol. 11, 123–144 (2017) <https://doi.org/10.1007/978-3-319-54024-5>
4. Ahmed, A., Islam, S., et al.: Big data analytics using neural networks for earlier cancer detection. *J. Med. Imaging Health Inf.* **7**(6), 1469–1474 (2017)
5. Liu, Y.: Incomplete big data imputation mining algorithm based on BP neural network. *J. Intell. Fuzzy Syst.* **37**(13), 1–10 (2019)
6. Wang, W.: Research of epidemic big data based on improved deep convolutional neural network. *Comput. Math. Methods Med.* **2020**(1), 1–10 (2020)
7. Kuai, W.: Prediction method for energy consumption of high-rise buildings based on artificial neural network and big data analysis. *NeuroQuantology* **16**(6), 524–530 (2018)
8. Li, C.Y., Zheng, L.: Analysis of Tai Chi ideological and political course in university based on big data and graph neural networks. *Sci. Program.* **2021**(1), 1–9 (2021)
9. Zhu, B., Chen, Y., Cai, Y., et al.: Network security situation prediction system based on neural network and big data. *Int. J. Secur. Its Appl.* **11**(1), 93–108 (2017)
10. Hao, Q., Qin, L.: The design of intelligent transportation video processing system in big data environment. *IEEE Access*, 1 (2020)
11. Ma, H., Pang, X.: Research and analysis of sport medical data processing algorithms based on deep learning and internet of things. *IEEE Access* **7**, 118839–118849 (2019)
12. Zhang, Z.: Big data analysis with artificial intelligence technology based on machine learning algorithm. *J. Intell. Fuzzy Syst.* **39**(5), 1–8 (2020)



Analysis of Financial Sharing Model Based on Block Chain Technology in Internet + 5G Environment

Yu Li^(✉)

Department of Management, Wuhan University of Technology, Wuhan, Hubei, China

Abstract. The advent of the Internet + 5G era has boosted the emergence and development of block chain technology. Applying it to the cost security control of the financial sharing model, it is found that block chain technology can reduce the company's redundant resources and strengthen the company's cost security management and control in four aspects: reducing the workload of financial sharing, simplifying the financial sharing process, improving information transparency and the security of financial sharing. For companies that have adopted or plan to adopt the financial sharing model, in the future, they can start with four aspects: establishing a sound financial sharing system, improving the level of computer technology support for financial sharing, upgrading the talent management system and improving the level of information security, so as to better realize the Internet + 5G environment. The cost security control of the financial sharing model of block chain technology improves the operational efficiency and efficiency of enterprises, and promotes the realization of healthy and sustainable development of enterprises.

Keywords: Internet + 5G · Block chain technology · Financial sharing · Cost security control · Information safety

1 Introduction

Since the first World Internet Conference was held 2014, mobile Internet technology has developed rapidly. In 2019, the Ministry of Industry and Information Technology officially issued a 5G commercial license, marking the official arrival of the 5G era. The 5th generation mobile communication network has faster speed, higher frequency band, more stable state, and lower network delay compared with 4G. Block chain technology is a distributed architecture system, which integrates a variety of computer technologies such as point-to-point transmission and encryption algorithms [1, 2]. It has the characteristics of decentralization, traceability, and high degree of openness. Network resources put forward very high requirements. The arrival of the Internet + 5G era has just met the needs of the block chain for storage computing resources, network speed and safety, and provided more reliable technical support and guarantee for the block chain.

In the process of development and growth of modern enterprises, they often establish a headquarters group with subsidiaries or branches. The establishment of branches often

establish the same functional management department as the headquarters and need to hire corresponding functional department staff. The repeated setting of the organization increases the human resource cost of the enterprise, and the information of the headquarters group and the subsidiary or branch cannot be synchronized and shared. In order to realize financial transformation, the establishment of a financial sharing center, simplifying the process, promoting the timeliness of accounting information transmission, improving work efficiency, and realizing the integration and sharing of financial information have become the general trend [3, 4]. Table 1 summarizes the relevant content of the policy guidelines related to information network security in recent years.

Table 1. The policy guidelines of informational network security

Issuing time	Policy guidelines	Related information
2013	Enterprise Accounting Information Work Standards	Enterprises should use modern information technology such as computers and network communications to carry out accounting, and use the above-mentioned technical means to organically integrate accounting with other business management activities
2016	China Block chain Technology and Application Development White Paper	Domestic block chain technology development road-map, relevant support policies, technical research, platform construction, and application demonstrations
2019	Regulations on the Management of Block chain Information Services	Block chain information service providers should implement the main responsibility of information content security management
2019	Guiding Opinions on Strengthening Industrial Internet Security Work	Promote the implementation of industrial Internet security responsibilities; Build an industrial Internet security management system; Improve enterprise industrial Internet security protection levels; Strengthen industrial Internet data security protection capabilities; Promote industrial Internet security Technological innovation and industrial development
2021	Three-year action plan for the development of new data centers (2021–2023)	Innovation-driven, industrial upgrading; green, low-carbon, safe and reliable

Then, under the circumstance of the continuous improvement of Internet computer technology, doing research on how it affects the cost security control under the financial sharing model, and how to use its advantages to realize the cost control method of the financial sharing center, for companies that have achieved financial sharing, and companies that plan to achieve financial transformation in the future have certain reference and guiding significance.

2 The Influence of Block Chain on the Cost Security Control of the Financial Sharing Model Under the Internet + 5G Environment

In general, under the Internet + 5G environment, financial sharing has an increasing demand for computer information technology, and block chain technology can just meet this demand. The impact of block chain on the cost security control of the financial sharing model is specifically manifested in four aspects: reducing the workload of financial sharing, simplifying the process of financial sharing, improving information transparency and improving the security of financial sharing, the logical framework diagram is as follows (Fig. 1).

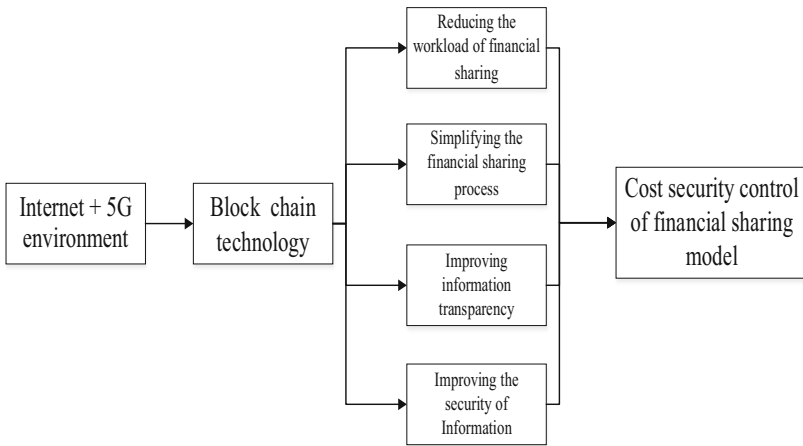


Fig. 1. The influence of block chain on the cost security control

2.1 Reducing the Workload of Financial Sharing

In the context of Internet + 5G, each information port of an enterprise group is connected as a whole. Each transaction of a subsidiary or branch is calculated, analyzed, and transmitted data according to the model of the block. The data of the same block can be grouped into one chain to form different systems, which can solve the dilemma of sharing financial data that is cumbersome and cannot be uploaded in large quantities. The

characteristics of smart contracts and consensus mechanisms in block chain technology ensure point-to-point automatic business transactions. Moreover, the block chain itself is a digital application technology, and the use of this technology only requires encryption key matching to realize the automated processing of most businesses. Some programs are handed over to computer technology to complete automatically, thus reducing the manual workload and reducing the management cost of the enterprise.

2.2 Simplifying the Process of Financial Sharing

Due to the high degree of automation of block chain technology, manual authorization steps are reduced, and the algorithm directly drives the authorization process, which reduces the manual authorization cost. In the past, financial work required the accounting staff to manually keep accounts based on the reimbursement of staff in various departments. However, in the Internet + 5G environment, due to the characteristics of fast internet speed and large storage capacity, the reimbursement personnel of each department, namely the bill maker, can use the financial sharing system in real time to upload clear images of the reimbursement documents according to the corresponding normative guidelines, and then pass the financial After the business reviewers of the sharing center have verified that they are correct, the acquiring post will receive the bills, and the financial sharing system will automatically generate accounting vouchers, so that the bill preparation process is simplified and the process cost is reduced.

2.3 Improving Information Transparency

All data recorded on the chain is open and transparent via using block chain technology. Each node is independent of each other and can be shared to access data information on the chain. With access rights, the staff can view the data of each subsystem in the financial sharing system. And any data modification or update in the shared system, whether it is the time when the data information changes, or the state before and after the change, will leave traces, which is also conducive to the process of tracing the entire data information in the future. Since the most prominent feature of the block chain is decentralization, financial sharing has achieved a certain degree of decentralization in this context, which further reflects the nature of sharing.

2.4 Improving the Security of Financial Sharing

The problems of financial sharing in the development process have become increasingly prominent, such as data leakage and data tampering. Since the data encryption algorithm in the block chain has the characteristics of one-way, irreversibility and non-tampering, and the Internet + 5G background provides a higher level of protection for the block chain technology, so it is applied technical services to the financial sharing system are more secure and reliable. Although personnel from various departments can access shared data under authorization, the continuous upgrading of block chain technology has also made corresponding improvements in user authentication and verification. Each individual section in the block chain is independent of each other, and the information in

each section can be traced back and is complete. Back-office staff can query the access traces in the system. If there is a security problem, they can find the corresponding source through the access traces and the corresponding time, and can monitor the company's books and transactions for real-time auditing [5]. This fully meets the requirements of the accounting standards for the security of financial sharing, reduces financial risks, and improves benefits.

3 The Method of Financial Sharing Cost Security Control Under the Internet + 5G Environment

After analyzing the influence of block chain on the cost security control of the financial sharing model, it is more important to better realize the security control of the financial sharing cost. Based on existing research, this paper summarizes the following four suggestions. Its logical framework is shown in Fig. 2.

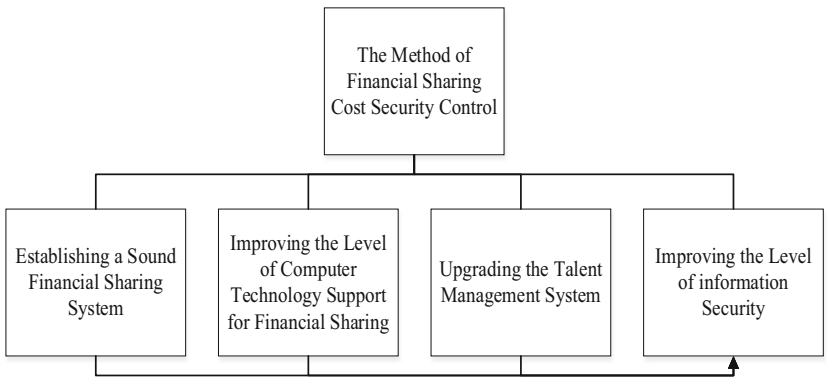


Fig. 2. The method of financial sharing cost security control

3.1 Establishing a Sound Financial Sharing System

Nothing can be accomplished without norms or standards. Only by taking information safety as the guide, establishing a strict financial sharing system and refining it to different business scenarios, can personnel in various departments have a reference system and norms in the process of realizing financial sharing. For employees who are unskilled in business operations, when they encounter unfamiliar business, they can refer to the clearly stipulated system, and they can follow the behavior and become familiar with the work under the guidance of the system, so as to gradually improve work efficiency. The performance appraisal and rewards and punishment system helps to build a fairer salary system in the context of Internet + 5G, and employees' efforts are rewarded accordingly, thereby increasing the stability of employees, reducing brain drain, and reducing human resource recruitment and training costs.

3.2 Improving the Level of Computer Technology Support for Financial Sharing

In the context of Internet + 5G, the speed of technological update and iteration is rapid, the update of the financial sharing system should also be synchronized with this. With reference to other relatively successful cases, in the process of pursuing standardization and unification, pay attention to the industry characteristics of the company's own industry, rely on the Internet + 5G environment, pay attention to the latest technological developments, and give full play to the block chain. The role of technology is to update and improve the existing information technology level in time, make full use of the ERP system, strengthen the automated management level of financial information, reduce the complexity of resources, realize the effective integration and allocation of resources, improve efficiency, and at the same time, pay attention to the risk management and control of the technology application process, strengthen risk awareness [6], and reduce the company's various resource costs.

3.3 Upgrading the Talent Management System

As a cutting-edge information technology, block chain has higher requirements for digital talents. Based on this, the content of employee job training is not limited to employee-related professional knowledge. In addition, it is necessary to integrate the big data concept of Internet + 5G into their work, so that they can understand and master the methods to realize the interconnection and intercommunication of data in various departments. Under the condition of no time and space constraints, through the company's internal financial sharing system, check and call information related to the work content, skilled in operating the information management platform. And in the process of using, put forward corresponding system optimization suggestions. This not only allows employees to improve their own business work capabilities, but also provides the most true and effective feedback for the company's system upgrades, paving the way for employees' work efficiency in the future, and reduces human resource costs.

3.4 Improving the Level of Information Security

Improving the level of information security needs to be based on a strict information confidentiality system, a mature technical support level and a high level of professional ethics of employees [7]. First, a strict information confidentiality system needs to be established, and the system is used as a reference for behavioral constraints. Staff who have the opportunity to access data and information must be authenticated multiple times, and the authentication process and standards are presented in the form of a clear text system [8]. If the company has established a risk management and control department, it can refine its responsibilities and requirements, work closely with the technical service department, and incorporate the work of maintaining information security. Secondly, in terms of technology, it is necessary to establish a sound virus defense system, to achieve daily maintenance and update of the system, to prevent hackers or competitors from attempting to obtain confidential company information through illegal means. At the same time, company must pay attention to the innovation of the company's internal

encryption technology. Finally, as the follower of the system and the user of technology, company employees should strengthen their professional ethics, establish a correct awareness of confidentiality, implement the information confidentiality system in place, and be proficient in the various technical requirements to improve the level of information security [9, 10].

4 Conclusion

With the support of the Internet + 5G background, block chain technology reduces the workload of financial sharing, simplifies the process of financial sharing, and improves information transparency and security of financial sharing. These four aspects have resulted in the cost security control of enterprises under the financial sharing model. For companies that have established and gradually improved the financial sharing model and those planning to implement the financial sharing model in the future, they can start from four aspects: establishing a sound financial sharing system, improving the level of computer technology support for financial sharing, upgrading the talent management system and improving the level of information security. In order to better realize the cost security control of the financial sharing model based on block chain technology under the Internet + 5G environment, improve the operational efficiency and efficiency of the enterprise, and help the enterprise achieve healthy and sustainable development.

References

1. Govender, S.G., Kritzinger, E., Loock, M.: A framework and tool for the assessment of information security risk, the reduction of information security cost and the sustainability of information security culture. *Pers. Ubiquit. Comput.* **25**(5), 927–940 (2021)
2. Yang, Y., Liu, Q., Song, J., Zhou, M.: The influence mechanism of financial shared service mode on the competitive advantage of enterprises from the perspective of organizational complexity: a force field analysis. *Int. J. Account. Inf. Syst.* **42**, 1–21 (2021)
3. Chen, X., Metawa, N.: Enterprise financial management information system based on cloud computing in big data environment. *J. Intell. Fuzzy Syst.* **39**(4), 5223–5232 (2020)
4. Song, Y., Zhang, F., Liu, C.: The risk of block chain financial market based on particle swarm optimization. *J. Comput. Appl. Math.* **370**, 112667 (2020)
5. Chen, Y., Zhang, Y., Zhou, B.: Research on the risk of block chain technology in Internet finance supported by wireless network. *EURASIP J. Wirel. Commun. Netw.* **2020**(1), 1–11 (2020)
6. Jiang, S.: Research on big data audit based on financial sharing service model using fuzzy AHP. *J. Intell. Fuzzy Syst.* **40**(4), 8237–8246 (2021)
7. Shaoqin, L.: Enterprise financial cost management platform based on FPGA and neural network. *Microprocess. Microsyst.* **80**, 1–5 (2021)
8. Arulananth, T.S., Baskar, M., Anbarasu, V., Thiagarajan, R., Rajendran, T., Balaji, A.: Multi party secure data access management in cloud using user centric block chain data encryption. *Pattern Recogn. Lett.* **152**, 295–301 (2021)
9. Nurul Asmui Azmi, M.A., Teoh, A.P., Vafaei-Zadeh, A., Hanifah, H.: Predicting information security culture among employees of telecommunication companies in an emerging market. *Inf. Comput. Secur.* **29**(5), 866–882 (2021)
10. Carmi, G., Bouhnik, D.S.: The effect of rational based beliefs and awareness on employee compliance with information security procedures: a case study of a financial corporation in Israel. *J. Inf. Knowl. Manag.* **15**, 109–125 (2020)



Performance Prediction Model Based on K-Means Clustering Algorithm

Wei Zheng, Lihong Hao, and Tao Liu^(✉)

Dadu River Hydropower Development Co., Ltd., Chengdu, Sichuan, China

Abstract. With the development of the times, more and more fields will gradually realize the objectivity and convenience of using data to describe the characteristics of characters. The essence of the key user profile of accurate services is data mining and natural language processing, of which the more widely used Including short text classification processing method, keyword extraction method, fusion pruning algorithm, TF-IDF algorithm and other methods. How to solve the problem of different data standards, complete ability and job matching, maximize employee value, achieve objective evaluation of employee performance, comprehensively record employee work behavior and abilities, scientifically improve company A's employee profile model, and use big data technology to manage human resources Empowerment is the core and key of A company's talent training and value drive. Through a large number of literature readings and experimental investigations, this paper gives a detailed overview of the performance model research under the flexible organizational structure of A company, the task-driven OKR of A company, and the establishment of the AI-based company A employee in-depth portrait platform. And use the K-Means clustering algorithm to establish a performance prediction model, and carry out experimental simulation and performance analysis. The K-means clustering algorithm is applied to the flow chart and experimental simulation and system performance test of the group connection point aggregation formation control. The experimental results show that the K-means clustering algorithm in the model process is applied to the flow chart of the cluster connection point aggregation formation control with accurate parameter data, which greatly improves the efficiency of human resource management.

Keywords: K-Means clustering algorithm · Performance prediction · Model research · Performance algorithm

1 Introduction

In recent years, with the continuous application and development of big data technology, domestic and foreign scholars have conducted multi-angle explorations on user portraits and human resource management models under flexible organizational structures [1]. Human resource management continues to develop towards digitization and intelligence [2]. The human resource management model under the flexible organizational structure has attracted the exploration and research of domestic and foreign experts and scholars [3]. In addition, scholars are still exploring the technology or other application scenarios

of user portraits. For example, Al-Janissary O I proposed to establish a K-Means clustering algorithm model [4]. The multi-dimensional label clustering proposed by Li W improves the service accuracy for learners' precise recommendation [5, 6].

2 Implementation Plan

2.1 Technical Route

- (1) Employee portraits of A company based on unsupervised machine algorithm [7, 8]. It is planned to use the "label + portrait" method to dig in-depth the business logic and data rationality of the data set, implement the employee behavior and ability tag library, and analyze and research the comprehensive characteristics of the employee, and then classify the attribute of the factor set, so as to classify the employee sample Process "fact tags", "model tags", and "advanced tags" to build "user portraits" of employee capabilities. Among them, the "fact label" represents the employee's own attribute data, the "model label" represents the result information of the employee in the sample distribution after statistical analysis, and the "advanced label" is based on the employee's past behavior and ability performance. The dimension makes future predictions to accumulate labels. By continuously accumulating the employee behavior and ability tag library, enriching and improving tags, we can accurately describe the overall characteristics of employees [9, 10].
- (2) The performance model under the flexible organizational structure of A company. It is planned to adopt the method of machine learning data modeling, use unsupervised machine learning algorithms to deeply explore the influence of behavior and ability factors on employee performance, and then build a set of performance prediction models that can be iterated and self-corrected in real time through a combined regression model. So as to regularly guide employees to self-examine and correct their work [11, 12]. A task-driven OKR research for company A. Through the setting and evaluation of key goals and key results, we can correctly understand the performance of employees in performing tasks. Through monthly, quarterly and annual feedback, we can accumulate and iterate the result labels in the label library to more objectively evaluate employees' performance. Contribution, ability growth, and sharing and help to other team members.
- (3) The establishment of an AI-based in-depth portrait platform for employees of A company. Based on the above model research on the content of human resource management of A employees, it is planned to build a micro-application platform to complete the overall UI design of the application platform based on the integration of the underlying data. Front-end design, and embedded research results in the platform, including multiple application modules, to build a deep portrait platform for A employees, and realize A's human resource management intelligent, precise, and agile.

2.2 Basic Conditions of Research Work and Project Implementation Methods

- (1) The research work of Performance forecast is carried out around critical paths such as demand analysis, data collection, data analysis and preprocessing, model construction, and reliability evaluation system. Comprehensive analysis of company A's corporate characteristics and employee composition, analysis of the differences in different types of employee abilities and behaviors, and objective construction of employee portrait business rules; research on employee portrait methods, through the establishment of a multi-level tag library, and unsupervised machine learning algorithms for sample aggregation Class, accurate description of employee characteristics; use natural language processing, including keyword extraction, referential disambiguation, part-of-speech tagging, etc. to complete task tag library establishment; research performance evaluation model under flexible organizational structure, select multiple logistic regression, decision tree, Etc. Various algorithms are compared, and the regression model of the optimal algorithm and parameters is fitted.
- (2) Performance forecast is highly professional and difficult in business. It is planned to adopt a joint development method to invite professionals to sort out business rules and design scenarios. At the same time, under the guidance of A company, the big data company assists the human resources department and is responsible for the project. Daily management and coordination rules. The implementation of the project is planned to be tendered and commissioned by units with professional advantages to give full play to the respective advantages of production enterprises and technical units to ensure the orderly and effective development of research topics. Before the implementation of the project, a detailed work outline was prepared according to the research tasks and strictly reviewed. During the implementation of the project, the assessment will be carried out in stages according to the specific assessment indicators determined in the work outline, and rolling management will be implemented. The project technical team is composed of the project leader, the subject leader, and the technical backbone, responsible for formulating the work outline and specific technical research and analysis to ensure the correctness of the research technical route and the level of the research results.
- (3) The data corresponding to the preliminary factor set comes from the company's human resources data platform, OA system, all-in-one card, 360 Tianjin terminal security management system, party building management system and other 12 existing information systems and questionnaire surveys. The data Rich and accessible. The machine learning algorithm to be adopted has been researched and applied to human resource management under the economic benefit flexible organization structure in the scene fields such as mail classification and customer portrait, and multiple application modules are established through employee portraits to enable intelligent performance assessment, ability and job matching, OKR, etc. Many functions are easily realized, so as to realize intelligent human resource management.

3 According to K-Means Clustering Algorithm

In order to deeply study the performance prediction model based on K-Means clustering algorithm, this paper uses K-Means clustering algorithm to establish a performance prediction model. The flow chart of K-means clustering algorithm applied to group connection point gathering formation control is a calculation process based on company employee performance data. This model algorithm can reduce the workload of human resources personnel, speed up performance prediction, and increase the workload of relevant personnel. The K-Means clustering algorithm is expressed as follows.

$$C(x, y) = \sqrt{(a_{x1} - b_{y1})^2 + (a_{x2} - b_{y2})^2} \quad (1)$$

$$n_p = \sum_{x=1}^m \frac{a_x}{m} \quad (2)$$

Where $x = (a_{x1} - b_{y2})$, $y = (a_{x1} - b_{y2})$ is two two-dimensional data objects, which respectively represent the location of the key point and the assembly point, N_p represents the assembly point of the p-Th group, m represents the number of key points in the p-Th group; $C(a_x - n_y)$ represents the m key points, respectively The minimum distance to p assembly points, and then m key points are divided into the same class as n_x . After the above process, the key points can be controlled to gather to the target key points, and the moving key point gathering behavior of the KGM model is completed.

4 Experimental Simulation and Performance Analysis

4.1 Experimental Background

The performance prediction model is based on a new type of engine. In each simulation stage, it will add some components to complete the addition of key simulation performance. The most important feature of the performance prediction model lies in the imitation of the movement of the connection points, the study of the correlation between the connection points, the realization of the routing algorithm and the data collection. The collection and analysis of news is done through reports and message processing tools. The association between connection points is related to their coordinates, communication radius and bit rate. Here routing is achieved by the relevant components deciding how to forward messages in the existing connection. The performance prediction model simulator not only has built-in the most typical routing protocols in the opportunistic network, but also has a passive routing module for data exchange with other simulation simulators.

5 Experimental Method

Figure 1 is a flow chart of K-means clustering algorithm applied to group connection point aggregation formation control. Researchers can enhance the performance of performance prediction through the relevant program development platform. The experimental results are mainly generated by the report module during the simulation. The report is collected. The reporting module receives events (for example, messages or connection events) from the simulation engine and generates results based on them. The result may be an event log processed by an external post-processing tool, or it may be a calculation summary/statistic in the simulator. The graphical user interface of the model shows a visual interface composed of the location of the connection points, the association between the connection points, and the data stored in the connection point cache.

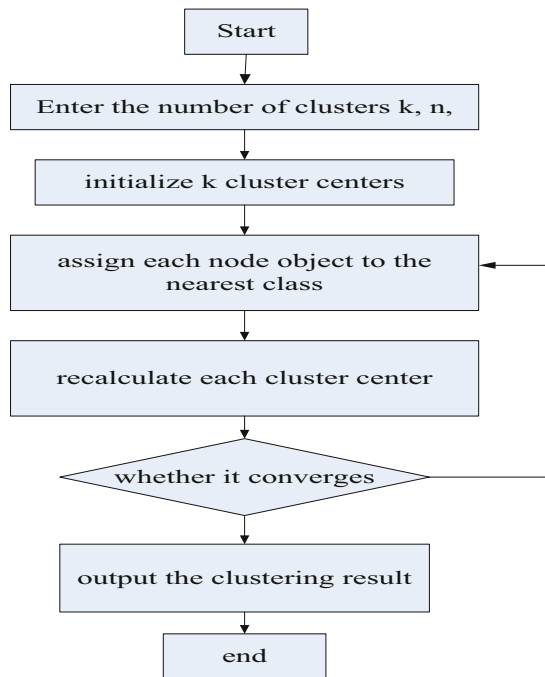


Fig. 1. The flow chart of K-means clustering algorithm applied to group node aggregation formation control

6 Experimental Simulation and System Performance Test

6.1 System Simulation Parameter Setting

The following is the specific simulation parameter table of the flow chart of K-means clustering algorithm applied to the group connection point aggregation formation control. The experimental data is shown in Table 1.

Table 1. Randomization model for privacy protection

Parameter type	Parameter value
Scene range (m ²)	4400 × 3300
Number of groups (a)	6
Communication range (m)	9
Node cache (M)	4
Node speed (m/s)	5.5
Routing Protocol	Epidemic
Simulation time (s)	4900
Number of nodes (a)	190
Free node (a)	0
Waiting time(s)	20–40
Aggregation start time (s)	1580
Dispersion start time (s)	3190
Transfer method	Bluetooth transmission

Table 1 is a flowchart of the specific simulation parameter table of the K-means clustering algorithm applied to the group connection point aggregation formation control. From the table, we can see the parameter types of the K-means clustering algorithm applied to the performance prediction model: The scene range is 4400 × 3300 (m²), the number of groups is 6 (units), the communication range is 9 (m), the node cache is 4 (M), the node speed is 5.5 (m/s), the routing protocol is Epidemic, and the simulation time is 4900 (s), number of nodes 190 (units), free node 0 (units), waiting time 20–40 (s), aggregation start time 1580 (s), dispersion start time 3190 (s), the transmission mode is Bluetooth transmission. It shows that K-means clustering algorithm is suitable for performance prediction model, its data is accurate, and it can improve the efficiency of human resources.

7 K-Means Clustering Algorithm is Used to Analyze the Type of Parameter Data in the Group Connection Point Aggregation Formation Control

As shown in Fig. 2, the experimental data in Table 1 is substituted into the K-means clustering algorithm to be applied to the flow chart of group connection point aggregation formation control (node number, node speed, node cache, communication range) accuracy and applicability the type of data analysis graph.

Figure 2 is the experimental data of Table 1 substituting the K-means clustering algorithm into the flow chart of the cluster connection point aggregation formation control (node number, node speed, node cache, communication range) accuracy and applicability data types diagram. The closer the applicability and accuracy are to 10, the

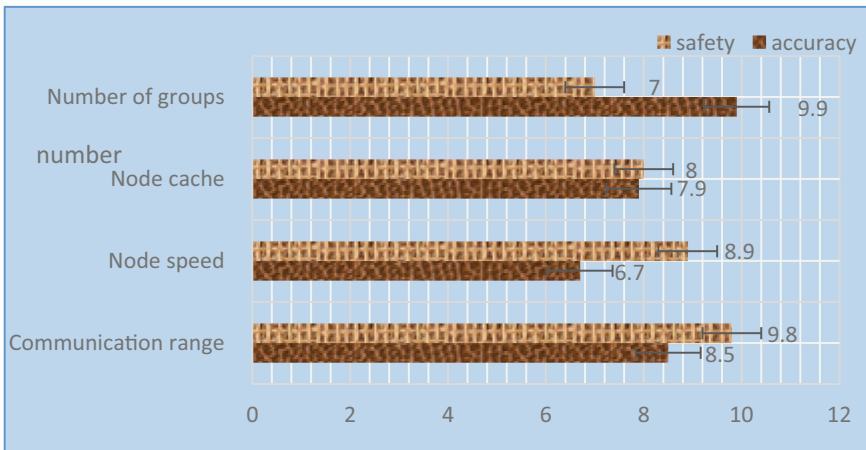


Fig. 2. Type analysis of parameter data in K-means clustering algorithm

better the applicability and accuracy of the model. It can be seen from the figure that the data accuracy of the number of nodes in the flow chart of the K-means clustering algorithm applied to the group connection point aggregation formation control is 8.5, the data accuracy of the node speed is 6.7, and the node cache data accuracy is 7.9, The data accuracy of the communication range is 9.9; the data applicability of the number of nodes in the flow chart of the K-means clustering algorithm applied to the group connection point aggregation formation control is 9.8, the data applicability of the node speed is 8.9, and the node cache data is applicable The sex is 8, and the data applicability of the communication range is 7. Test that all the above data are greater than 6, close to 10, indicating that the model has good applicability and high accuracy. Using this model to predict performance improves the efficiency of human resource management.

8 Conclusions

With the deepening of the construction of smart enterprises, the company is showing a trend of flexible management, and the organizational form is gradually becoming flexible. The work relationship network of employees is complex, there are many special tasks and temporary assignments, and cross-departmental collaboration is involved. This puts forward more for human resource management. Sophisticated and smarter development requirements. This paper studies the employee portraits of A company based on unsupervised machine algorithms, the performance model under the flexible organizational structure of A company, the task-driven OKR research of A company, and the establishment of an AI-based company A employee in-depth portrait platform. And use the K-Means clustering algorithm to establish a performance prediction model, and carry out experimental simulation and performance analysis. The K-means clustering algorithm is applied to the flow chart and experimental simulation and system performance test of the group connection point aggregation formation control. And the type analysis of the parameter data of the K-means clustering algorithm in the group connection point aggregation formation control. Design a set of solutions for intelligently matching employees

and tasks under a flexible organizational structure. The experimental results show that the K-means clustering algorithm in the model process is applied to the flow chart of the cluster connection point aggregation formation control with accurate parameter data, which greatly improves the efficiency of human resource management.

References

1. Qin, X., Li, J., Hu, W., et al.: Machine learning k-means clustering algorithm for interpolative separable density fitting to accelerate hybrid functional calculations with numerical atomic orbitals. *J. Phys. Chem. A* **124**(48), 10066–10074 (2020)
2. Gong, W., Pang, L., Wang, J., et al.: A Social-aware K means clustering algorithm for D2D multicast communication under SDN architecture. *AEU-Int. J. Electron. C.* **132**(2), 153610–153611 (2021)
3. Al-Sanjary, O.I., Roslan, M., Helmi, R., et al.: Comparison and detection analysis of network traffic datasets using k-means clustering algorithm. *J. Inf. Knowl. Manag.* **3**, 2050026–2050027 (2020)
4. Li, W., Zhao, J.Y., Yan, T.S.: Improved K-means clustering algorithm optimizing initial clustering centers based on average difference degree. *Kongzhi yu Juece/Control and Decision* **32**(4), 759–762 (2017)
5. Xia, D., Ning, F., He, W.: Research on parallel adaptive canopy-k-means clustering algorithm for big data mining based on cloud platform. *J. Grid Comput.* **18**(2), 263–273 (2020). <https://doi.org/10.1007/s10723-019-09504-z>
6. Ma, J., He, J., Shi, J., et al.: Nonlinear compensation based on k-means clustering algorithm for nyquist PAM-4 VLC system. *IEEE Photon. Technol. Lett.* **31**(12), 935–938 (2019)
7. Wang, X., Zhang, Q., Xin, X., et al.: Robust weighted K-means clustering algorithm for a probabilistic-shaped 64QAM coherent optical communication system. *Opt. Express* **27**(26), 37601–37613 (2019)
8. Wu, H., He, Y., Zhao, B., et al.: Research on dynamic equivalent of wind farm based on improved k-means clustering algorithm. *Taiyangneng Xuebao/Acta Energetica Solaris Sinica* **39**(11), 3232–3238 (2018)
9. Boue, A., Lesage, P., Cortes, G., et al.: Performance of the ‘material failure forecast method’ in real-time situations: a Bayesian approach applied on effusive and explosive eruptions. *J. Volcanol. Geothermal Res.* **327**, 622–633 (2018)
10. Rohmer, J., Chojnacki, E.: Forecast of environment systems using expert judgements: performance comparison between the possibilistic and the classical model. *Environ. Syst. Deci.* **41**(1), 131–146 (2021). <https://doi.org/10.1007/s10669-020-09794-9>
11. Dandl, F., Hyland, M., Bogenberger, K., Mahmassani, H.S.: Evaluating the impact of spatio-temporal demand forecast aggregation on the operational performance of shared autonomous mobility fleets. *Transportation* **46**(6), 1975–1996 (2019). <https://doi.org/10.1007/s11116-019-10007-9>
12. Dai, G., Mu, M., Jiang, Z.: Evaluation of the forecast performance for North Atlantic oscillation onset. *Adv. Atmos. Sci.* **36**(7), 753–765 (2019). <https://doi.org/10.1007/s00376-019-8277-9>



The Design of Contemporary Urban Literature System Based on Internet of Things Data Processing Technology

Qiong Zhong^(✉)

College of General Education, Guangdong University of Science and Technology,
Dongguan 523083, China

Abstract. With the continuous advancement of my country's urbanization process, the role of literature in urban economic, political, and social development has become increasingly prominent. As a special kind of labor, urban literature contains the essential stipulations of human beings, which embodies the unity of the objective world and the subjective world in the transformation of the urban subject. With the development of the Internet of Things era, data processing technology continues to improve. Based on this, this paper uses the powerful data processing advantages of computer networks to design a contemporary urban literature system, conducts data processing on related urban literature research documents in the system, and it is found that urban literature covers a wide field and a large span, which promotes the construction of urban culture.

Keywords: Internet of Things · Data processing technology · Contemporary urban literature · System design

1 Introduction

Urban literature is the wealth of people who gather material and spiritual culture in long-term urban life. As far as urban development is concerned, urban literature, as a social material carrier, promotes the continuous development of human society and creates a cultural scene with urban characteristics. Therefore, this paper studies the design and application of contemporary urban literature system to provide basic equipment and conditions for strengthening the construction of urban literature.

Many scholars have studied the contemporary urban literature system design based on the data processing technology of the Internet of Things, and have obtained rich research results. For example, a scholar believes that urban literature is the driving force of urban development, marking the level of urban development, and also reflecting the image of a city. Urban literature can attract talents, technology, and capital, and form the core competitiveness of urban literature, which is conducive to promotion. The construction of urban literature promotes the sustainable development of the city [1]. A researcher used the powerful computing and storage information functions of the Internet of Things to integrate predecessors' research data on contemporary urban literature, and created

an urban literature system with data processing capabilities. As long as people use the system, the system will record the user's browsing traces, and then form urban literary recommendations according to user preferences, and stimulate people's sense of literary participation [2]. Although the research results of contemporary urban literature system based on Internet of Things data processing technology are good, compared with mature urban development concepts, there is still no in-depth research. There is a long way to go in the construction of urban literature influencing urban development.

This article introduces the data processing methods of the Internet of Things and the significance of contemporary urban literature construction. Aiming at the design principles of the urban literature system, the contemporary urban literature system architecture is designed using the Internet of Things data processing technology. The data collection function is set on the system, and then the system is applied to In the quantitative analysis of urban literature research literature, the practicability of data processing technology is verified.

2 Internet of Things Data Processing and Contemporary Urban Literature Construction

2.1 Internet of Things Data Processing Methods

(1) Wireless sensor network data processing

The characteristics of wireless sensor networks that are different from traditional Internet networks have brought many new difficulties to its data processing. Compared with traditional data, WSNs data has the following characteristics: First, real-time and heterogeneous. IOT terminal devices are generating high-speed data streams all the time. Many IOT-based applications also have high requirements for real-time data processing. Therefore, IOT data processing must fully tap the computing resources of the terminal nodes to achieve intelligent real-time perception and precise control. The second is semantic richness. The event stream of the Internet of Things not only carries information about time and space, but also implies the relationship between context and upper-level applications, making necessary preparations for constructing complex events with rich meaning. The third is limited energy. The sensor node is a small built-in device, which inevitably requires consideration of storage space, computing speed, and power energy limitations when processing data to extend the life of the node and the network cycle. The fourth is uncertainty. Due to the influence of factors such as the accuracy of the IOT hardware devices and the external environment, the data read often appears to be read more, missed, read repeatedly, and read dirty data [3, 4].

(2) RFID data processing

Radio frequency identification is a technology that uses micro-signals to identify objects and obtain object-related data and information that can be identified spontaneously without contact. RFID data has the characteristics of continuity, correlation, inaccuracy, heterogeneity, and semantic richness. It is very similar to the event stream. The only difference is that its data comes from each reader terminal. It is difficult but very meaningful to discover and extract regular and complex events from these large amounts of raw data [5].

- (3) IOT data processing based on cloud computing
 “Internet of things” and “cloud computing” are two inseparable concepts. People propose to introduce cloud computing into the application of the Internet of Things, and make full use of the large-scale parallel computing capabilities of cloud computing and massive data centers to complete Real-time, heterogeneous, massive IOT data calculation. Data processing of the Internet of Things through cloud computing has become one of the most commonly used methods. SCOPE is a parallel processing language developed by Microsoft for the analysis of massive data in a cloud computing environment. It adopts a relational data model, inherits some of the characteristics of SQL, and allows users to define their own operators. BigTable is a distributed multi-dimensional mapping table, the structure of specific data processed by users. But BigTable does not provide APIs that support all relational database functions, and it has abandoned the atomicity requirements of relational databases. In addition, it only supports simple read, write and storage functions, and does not support transaction processing in the traditional sense [6].

2.2 Basic Principles of Contemporary Urban Literature System Design

- (1) Ease of use
 The development of the upper computer management system for data adopts the B/S structure mode. Users can run the application directly in the browser without installing any plug-ins, which is convenient for project implementation and equipment deployment, as well as for its management. The design of the user operation interface is as friendly and user-friendly as possible, to the greatest extent possible for users to understand the system and to simplify the operation process as much as possible. In the case of user misoperation, prompt friendly prompts to it. The handheld part of the system adopts a friendly interface design, which is convenient for users to input and query [7].
- (2) Stability
 Reliability is the primary condition for the system to be applied. The built system must not only be able to operate stably, but also have strong fault tolerance and a certain degree of resilience after failure. In the case of abnormal system operation, protect the important data of the system as quickly as possible, and send alarm reminders to the management personnel to minimize the loss caused by the abnormal situation.
- (3) Security
 Security is an important indicator of modern intelligent systems. The system uses different role management and different access authority settings to protect the security of important data as much as possible. The system should have a recording function, which can save the user’s login behavior and query records, so that it can be dealt with in time when problems occur. In addition to ensuring the security of data, the application system needs to realize mass storage of a large amount of data, using cloud storage to ensure the security of data storage [8].

(4) Openness

System construction also needs to consider the principle of openness. The needs of users are diverse and ever-changing. When the system needs to expand business and other needs, the system needs to be improved and upgraded, so open design is carried out in the system development and design stage. It is necessary to ensure the integrity of the system, especially the interface part must be complete, while satisfying that the system has good scalability and is convenient for system upgrades [9].

2.3 The Significance of Contemporary Urban Literature Construction

(1) Promote the all-round development of people in the city

The continuous improvement of citizens' cultural needs is the key to the progress of urban cultural construction. With the improvement of my country's scientific and technological production level, the economic strength has increased significantly, and the material life of urban residents has become increasingly richer, which has driven the development of the demand for knowledge and cultural products. Therefore, it is necessary to strengthen the construction of urban culture, improve the urban cultural industry system, continue to introduce new cultural products and services, and meet the growing spiritual and cultural needs of citizens, so that citizens can continuously improve and develop in cultural consumption. This is also the construction of urban culture today. One of the important topics of [10, 11].

(2) Urban cultural construction promotes urban economic development

With the advent of the era of knowledge economy, people have consciously realized that culture exists as another value that is relatively independent of the economy. The nature of cultural value has also been re-examined. As urban economic development gradually shifts from relying on tangible material resources to knowledge, the importance of urban culture in promoting urban development has become more obvious. In the era of knowledge economy, the promotion and integration of urban economy and urban culture is the general trend. The interaction between urban economy and urban culture is reflected in the development of urban economy, which promotes the prosperity of modern urban culture. The prosperity of urban culture provides a newer and stronger driving force for the further development of urban economy. In our country, influenced by the traditional individual economic values, they only one-sidedly emphasize the rapid economic growth, thus neglecting cultural construction and the creative value of cultural resources and cultural productivity, so that there are still many cities that have not formed positive and characteristic cultural concepts [12].

(3) Realize sustainable urban development

In the ever-developing era of economic globalization, the information cycle has slowed down significantly, and the density of information exchange has exploded. Only a learning city can rely on its strong learning ability and adaptability to calmly cope with complex and changeable situations, and reflect the requirements of urban

scientific development for openness, tolerance and continuous progress in different realistic cultural environments.

2.4 Data Fusion Algorithm Based on the Internet of Things

The fusion function is mainly determined by fusing the data with complex algorithms in the mean, median, maximum or minimum values. Assuming that there are n sensor nodes in a multi-sensor data fusion system, and their output data are X_1, X_2, \dots, X_n respectively, then the fusion function of the system can be expressed as:

$$D(x_1, x_2, \dots, x_n) = y \quad (1)$$

In formula (1), D represents the fusion function, and y represents the result of the data fusion of these n nodes.

When multiple data are waiting to be merged, the tolerance function is defined as follows:

$$\begin{aligned} H : X * X * \dots * X &\rightarrow [0, 1] \\ H(x_1, x_2, \dots, x_n) &= \min_{i,j} \{H(x_i, x_j)\} \end{aligned} \quad (2)$$

In formula (2), H represents the tolerance function, the value of the specified tolerance function is in the interval $[0,1]$, $H(x_i, x_j)$ represents the tolerance result of two sensor nodes, and $H(x_1, x_2, \dots, x_n)$ represents n sensors overall tolerance result of the node data.

3 Contemporary Urban Literature System Design Based on Internet of Things Data Processing Technology

3.1 System Architecture Design

Figure 1 shows the contemporary urban literature system architecture based on the Internet of Things data processing technology. The system includes four modules, namely, a user management module, a query statistics module, a system maintenance module, and a real-time monitoring module. The system maintenance module also includes database initialization, data backup and recovery, and basic data maintenance.

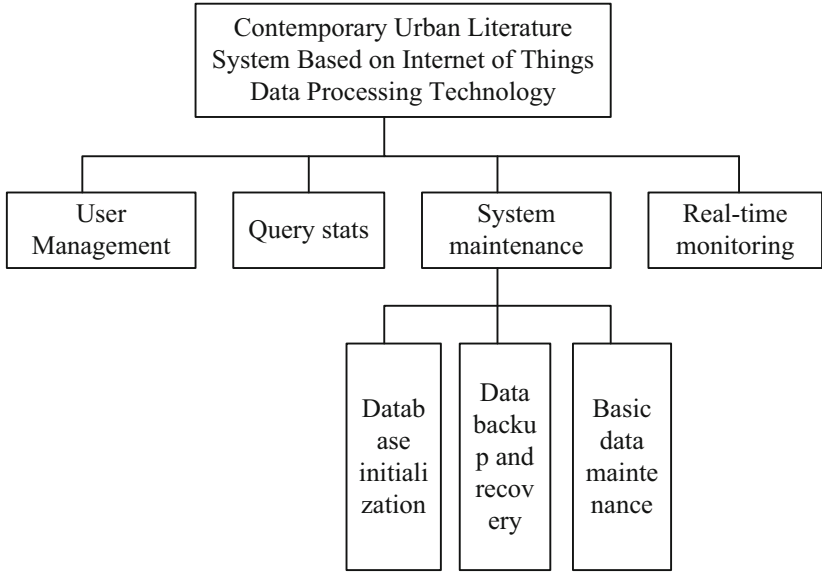


Fig. 1. Contemporary urban literature system architecture

3.2 System Data Collection Function Design

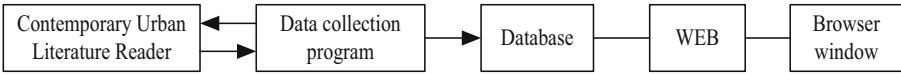


Fig. 2. Data collection diagram of contemporary literature system

Figure 2 shows a schematic diagram of collecting contemporary literature system data using IOT data processing technology. First, query contemporary urban literature information in the browser window based on WEB, secondly collect relevant information on the contemporary urban literature reader through a data collection program, and finally store the inquired and collected information in the literature data information database.

4 Application of Contemporary Urban Literature System Based on Internet of Things Data Processing Technology

4.1 Quantitative Analysis of Contemporary Urban Literature Research Literature

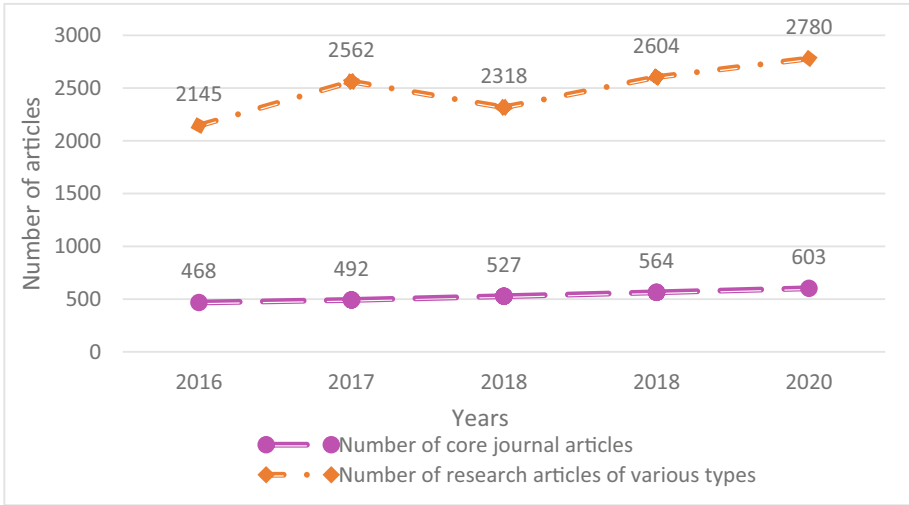


Fig. 3. The number of contemporary urban literature research literature

Figure 3 shows the amount of urban literature research literature analyzed in recent years using the Internet of Things data processing technology when applying the contemporary urban literature system. According to the data in the table, in 2016–2020, the number of research articles in core journals of contemporary urban literature has maintained a trend of increasing year by year, from 468 articles in 2016 to 603 articles in 2020. However, the number of articles related to contemporary literature research fluctuates greatly. Between 2016 and 2018, it first showed an increasing trend, then showed a downward trend, and then maintained an upward trend. Through the data processing function of the Internet of Things, we can understand the research situation of contemporary literary literature. From the research results, the number of contemporary urban literary articles has increased every year, which has a positive role in promoting the construction of contemporary literary city theory in my country.

4.2 Discipline Analysis of Contemporary Urban Literature Research

There are dozens of subject categories on urban urban literary design in my country. The research of this article is to use the Internet of Things technology to analyze the data of the eight subjects involved in the most. As shown in Table 1, the number of related articles on contemporary urban literature in 2020 is divided by subject category. Figure 2

Table 1. Subject categories involved in contemporary urban literature

Subject category	Number of articles	Percentage (%)
Building science	428	15.40
Drama film and TV literature	236	8.49
Sociology and statistics	331	11.91
Tourism	98	3.53
Art, calligraphy	205	7.37
Sculpture and photography	113	4.06
Music dance	97	3.49
Literary theory	74	2.66
Other	1198	43.09

shows that the total number of articles on contemporary urban literature in 2020 is 2,780, and there are 428 architectural science articles, accounting for 15.40% of the total., 236 articles on drama, film and television art, accounting for 8.49% of the total, 331 articles on sociology and statistics, accounting for 11.91% of the total, and 98 articles on tourism. It accounts for 3.53% of the total. There are 205 articles on art and calligraphy, accounting for 7.37% of the word count, 113 articles on sculpture and photography, accounting for 4.06% of the total, and 97 articles on music and dance, accounting for the total. 3.49%. There are 74 articles related to literary theory, accounting for 2.66% of the word count. The remaining 1198 articles are in subjects other than these eight subjects, accounting for 43.09% of the total number of articles. That is to say, these eight articles are related to contemporary urban literature research. The number of articles in the subject accounted for more than half of the total number of articles. Through this analysis, we can see the influence and interdisciplinary nature of urban literature in other disciplines.

5 Conclusion

This paper designs a contemporary urban literature system, and analyzes the number and subject types of relevant contemporary urban literature in the system based on the data processing technology of the Internet of Things. The analysis results reflect that the number of documents related to contemporary urban literature has increased in recent years, indicating that the number of people studying urban literature is increasing and the construction of urban culture has been accelerated; in addition, contemporary urban literature involves many types of disciplines, reflecting urban culture diversity.

References

1. Turunen, T.V., Eloranta, Hakanen, E.: Contemporary perspectives on the strategic role of information in internet of things-driven industrial services. *J. Bus. Ind. Market.* **33**(6), 837–845 (2018)
2. Xu, W., Zhang, J., Huang, S., et al.: Key generation for Internet of Things: a contemporary survey. *ACM Comput. Surv.* **54**(1), 1–37 (2021)
3. Shibuo, Y., Kawaika, K., Furumai, H.: Special issue on actions toward futuristic urban flood risk research and management. *J. Dis. Res.* **16**(3), 309–309 (2021). <https://doi.org/10.20965/jdr.2021.p0309>
4. Mohamed, S., Mohamed, D.: Data in the Internet of Things era: the propretization of data in light of contemporary business practices. *Int. J. Bus. Soci.* **21**(1), 81–94 (2020)
5. Wang, J., Guo, Y., Jia, Y., et al.: Modeling and application of the underground emergency hedging system based on Internet of Things technology. *IEEE Access* **99**, 1 (2019)
6. Henwood, M.: The potential of the li and the wangcheng for contemporary urban design. *Urban Morphol.* **22**(2), 159–161 (2018)
7. Galdini, R.: Urban re-use practices in contemporary cities: experiences in Europe. *Cities* **87**, 103–105 (2019)
8. Si, H., Sun, C., Li, Y., Qiao, H., Shi, L.: IoT information sharing security mechanism based on blockchain technology. *Future Gen. Comput. Syst.* **101**, 1028–1040 (2019). <https://doi.org/10.1016/j.future.2019.07.036>
9. Manogaran, G., Lopez, D., Chilamkurti, N.: In-Mapper combiner based MapReduce algorithm for processing of big climate data. *Future Gen. Comput. Syst.* **86**, 433–445 (2018). <https://doi.org/10.1016/j.future.2018.02.048>
10. Bhatt, A.P., Sharma, A.: Quantum cryptography for Internet of Things security a. *J. Electron. Sci. Technol.* **17**(3), 213–220 (2019)
11. Al-Rubaye, S., Kadhum, E., Ni, Q., et al.: Industrial Internet of Things driven by SDN platform for smart grid resiliency. *Internet of Things J. IEEE* **6**(1), 267–277 (2019)
12. Bensaali, F., Zhai, X., Amira, A., et al.: Guest editorial special issue on real-time data processing for Internet of Things. *IEEE Internet of Things J.* **5**(5), 3487–3490 (2018)



Laboratory Information System Under Big Data Technology

Xianjin Lin^(✉) and Ziyu Zheng

Development of IT Development and Management, Huaqiao University,
Quanzhou, Fujian, China
linxj@hqu.edu.cn

Abstract. Building laboratory information systems and improving business operations and service levels have become the only way to solve industry problems and clear obstacles to business development and industry progress. This article aims to study a laboratory information system based on big data technology. Through the understanding and research of the Internet of Things, we have designed and implemented a laboratory information system based on the Internet of Things, strengthened the management and control of laboratory machinery and equipment business, and improved the use of communication data of machinery and equipment in colleges and universities, as teaching equipment Business promotion provides technical support to verify the effect of laboratory information system optimization. This work uses big data technology to analyze the laboratory information system database under three physical server test scenarios: the virtual server before optimization and the virtual server after optimization. Performance. The experimental results show that the higher TPS is increased by 31% compared with before optimization, and by 25% compared with the test physical server.

Keywords: Big data technology · Laboratory information · System analysis · Equipment monitoring

1 Introduction

At present, there are a large number of classrooms in colleges and universities, some classrooms have a high utilization rate, and some have a low utilization rate, which cannot achieve an effective balance [1, 2]. There is no relevant system to collect the status, use efficiency and use cycle of multimedia equipment in the classroom, so that it is impossible to provide effective opinions on the update and replacement of experimental equipment [3, 4]. At present, there are often situations in which there are people with air conditioners and fans in the classrooms, but the Internet of Things is a new turning point in the development of information technology in today's world [5]. Custom transaction methods may not be compatible with the new transaction features provided by Internet objects. Therefore, we urgently need to formulate new development strategies to meet the needs of the new situation and the development of business [6, 7].

At present, some scholars have conducted preliminary research on big data technology [8]. Bo WA compares and analyzes the status quo, process, deficiencies, and

information collection of different service consulting systems, and builds an intelligent collection model of service consulting system information. This model integrates the basic process of consulting information collection, using data mining technology to realize the integrated functions of service consulting information collection, processing, and intelligent analysis, enhancing the collaboration and intelligence of information and intelligence collection, and realizing service consulting Automation and intelligence of intelligence [9]. Mo N combined with the current powerful cloud computing technology for data analysis, designed and analyzed a farmland planning area information collection system. The data control framework and theoretical structure model of farmland structure and regional information collection system are established. The software control process and hardware configuration of the acquisition system were gradually optimized, and a relatively complete farmland planning regional structure regional information acquisition system was obtained [10]. Research on the feasibility of laboratory information systems under big data technology.

This article first introduces the development and importance of the program, as well as the current status of the research and development of the Internet of Things building materials at home and abroad. Based on the Internet of Things technology, big data technology, cloud computing technology, etc., through a variety of intelligent detection technologies, real-time tracking and acquisition of personnel, equipment, resources and other information, enabling administrators to improve the classroom concept through the big data display platform, and quickly and accurately obtain information, Intelligent prediction, intelligent conclusion, improve the level of classroom management and teaching management, and provide relevant decision-making basis [1]. Introduced system solutions, system architecture technology, Internet of Things technology and data collection technology. Explain system software and application architecture, IoT communication solutions, business system design and design, and finally explain data design and optimization. Explain software testing methods, design test cases for system testing, and explain how to complete system testing.

2 Research on Laboratory Information System under Big Data Technology

2.1 Big Data Analysis

Traditional data analysis has changed from single site analysis to general content analysis. Knowledge acquisition, value exploration and relationship search are the key elements of major media information analysis. If traditional data analysis is to find gold from mines, the choice is important, but big data analysis is to pan for gold and store real gold [11].

One of the main characteristics of Hadoop is its size in the database and storage, which can lead to the ability to edit and add more clusters. It also saves a lot of money because it does not require high-tech servers. Big data analysis aims to use the Hadoop file distribution system (HDFS) as a repository. Sometimes, analysis tools use the services provided by the Hadoop system to process data. In terms of price and performance, it is difficult to find another good Hadoop solution. MapReduce is a platform that uses a large

number of computers (nodes) (collectively referred to as clusters or grids) to develop solutions to complex problems.

2.2 Combination of Internet of Things and Cloud Computing

The scale and understanding of the Internet of Things requires data collection and processing, and the characteristics of cloud computing, high power consumption, multiple users, high reliability, and large capacity are exactly what the Internet of Things needs. Develop scale and understanding. Cloud computing promotes the distribution, speed analysis and optimization of IoT sensor background data through large-scale computing clusters and high-speed transmission capabilities; powerful cloud computing technologies, including server tracking, network synchronization, and storage capacity. Modern technology makes it easier to create more applications; the high reliability and high degree of cloud computing provide highly reliable services for the Internet of Things. Maintaining these services requires computing software to be very reliable and easy to expand, which allows cloud computing to provide support services for Internet objects.

2.3 Platform Design Goals

2.3.1 Improve Classroom Utilization and Coordinate Management

Through classroom usage, we can clearly understand the school's classroom usage rate and idle rate. It is convenient for school administrators to understand the schedule of each teaching building. It is convenient for the school to make further arrangements for the construction of smart campuses such as site expansion and classroom upgrades according to the usage situation.

2.3.2 Equipment Monitoring, Early Warning

Through big data analysis, real-time control and dynamic monitoring of equipment distribution and usage can be achieved. Early warning reminders of equipment damage, failures, and equipment that need to be scrapped are convenient for management personnel to quickly deal with them in advance, and prevent failures before they occur.

2.3.3 Scientific Supervision to Promote the Improvement of the Management Level of Evaluation

Through scientific and objective big data statistical analysis, improve the management level of supervision and evaluation, free the teacher from the tedious task of supervision and evaluation, and focus on teaching work.

3 Investigation and Research of Laboratory Information System Under Big Data Technology

3.1 System Construction

Laboratory information system can be divided into four subsystems: preprocessing information system, big data analysis system, laboratory information management system

and user-centric system. The preliminary data collection system includes data collection information, other data collection information and data release activities; big data analysis system, specifically HDFS data storage, analysis and MapReduce prediction; data extraction, analysis and summary can be realized according to different campuses and different schools. Data display of the teaching building situation, real-time display including classroom status, equipment status, energy consumption statistics, environmental data, operation and maintenance overview, supervision and evaluation and other big data function information.

3.2 Performance Optimization

In this article, virtual non-uniform memory access architecture and virtual symmetric multi-processor architecture are used to increase CPU shared cache usage and improve physical memory access speed. By changing the detection frequency of the workload size and modifying the memory reservation of the virtual machine, the memory efficiency of the virtual machine is effectively improved. Changing the virtual disk type, adding a vSCSI controller, and switching to the virtual VMware type effectively improves the I/O performance of the virtual disk, allowing shared storage even under high CPU load and I/O. Will not become the bottleneck of the virtual machine server.

3.3 Design of Joint Kalman Filter

Taking into account the shortcomings of traditional GPS transmission systems in controlling the state of experimental equipment in colleges and universities, traffic signals in dense areas will be blocked and lost. The GPS positioning system combined with the operation algorithm of the WSN positioning system, simulation experiments prove that the average positioning method can improve the stability and accuracy of the system positioning. The Kalman Institute calculates the system by determining the variance of statistical errors. Generally speaking, process measurement must be described by a linear system. Both the received signal strength and mobile users can be approximated as a single-line system. The line system can be described by two equations:

The process equation is shown in Eq. 1:

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}, p(w) \sim N(0, Q) \quad (1)$$

The measurement equation is shown in Eq. 2:

$$z_k = Hx_k + v_k, p(v) \sim N(0, R) \quad (2)$$

x_k and z_k are the estimated value and measured value of the system at time k , respectively. u is the known system input.

3.4 Test Process

In the testing process of this article, all user databases and loads are created and simulated by BenchmarkFactory. For the current major databases, their size is about 100 GB. For the test database created by BenchmarkFactory, each increase in Scale (scale parameter)

increases by approximately 9.74 GB. Therefore, this article sets the scale to 10 to allow BenchmarkFactory to create a 97.4 GB test database. UserLoad (user load parameter) specifies the range of the number of virtual users accessing the database at one time from 100 to 2000.

4 Investigation and Analysis of Laboratory Information System under Big Data Technology

4.1 System Function Module Realization

The realization of the system function modules in the cloud computing operation management platform of the construction machinery Internet of Things includes six parts: laboratory status management, laboratory equipment status, energy consumption statistical data management, environmental overview, operation and maintenance overview, and supervision and evaluation.

The laboratory status management module is mainly used to include basic classroom data, classroom usage, teaching building scheduling rate, teaching building class scheduling, teaching building hotspot distribution, classroom scheduling statistics, classroom heat statistics, classroom scheduling data abnormal warnings, etc.

Users inquire about the basic information of construction machinery through the machinery management function, including name, equipment unique identification number, equipment number, machine type, unit and region, etc., to quickly understand the basic conditions of construction machinery. The laboratory state management part is implemented by SpringBoot framework, and the code structure is shown in Fig. 1.

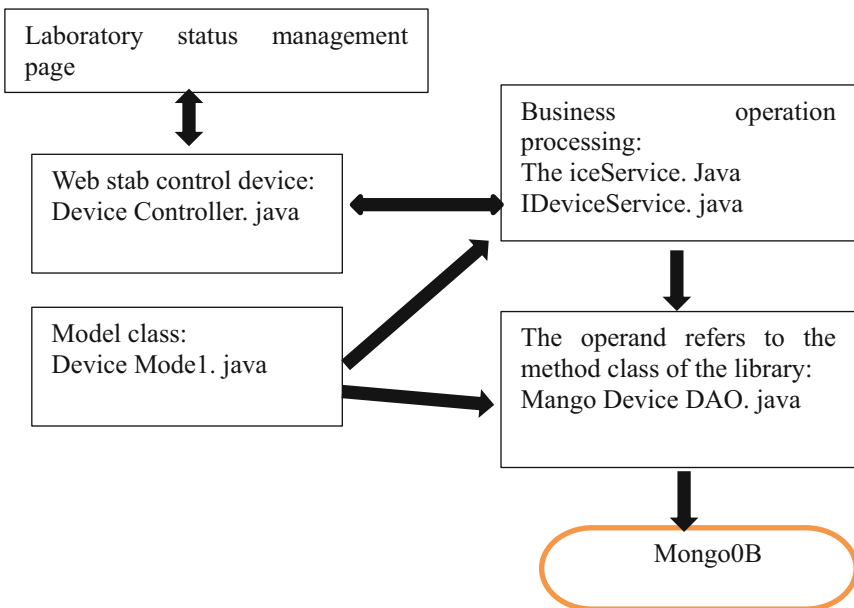


Fig. 1. Vehicle management code structure diagram

The experimental equipment status module includes early warning information, equipment usage status, statistics on the number and proportion of equipment opening abnormalities, equipment status of each teaching building, equipment communication status of each teaching building, equipment utilization rate, equipment failure rate, etc. Click the view icon in the “Operation” of a fault information to view the basic information of the current fault. Click the confirmation in a fault information operation to confirm the basic information of the current fault.

Energy consumption statistics management includes energy consumption statistics, accumulated expenses, voltage thresholds, meter device status, energy consumption of each classroom, energy consumption details of a single classroom, energy consumption history graphs, energy consumption trend graphs, abnormal data warnings, etc.

The environmental profile module includes classroom vacancy ranking, classroom air quality overview, single classroom air quality trend chart, early warning data, etc.

The operation and maintenance overview module includes equipment overview, purchase time statistics, repair report statistics, total equipment running time, supplier maintenance efficiency statistics, equipment alarm statistics, etc.

The supervisory evaluation module includes the types of evaluations, fast due plans, evaluation rankings, teacher qualification rates, course qualification rates, evaluation rates, and lesson evaluation plans.

4.2 Analysis of Performance Test Results

This test compares the performance of the physical server, the performance of the virtual server before optimization, and the performance of the optimized virtual server. The results are shown in Fig. 2. It can be seen from the figure that when User Load ≤ 1000 ,

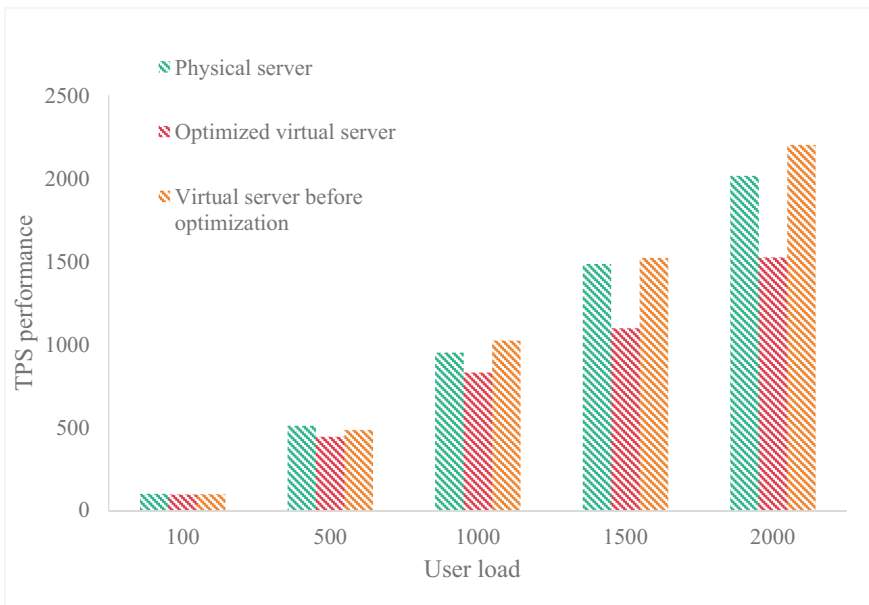


Fig. 2. Performance comparison between the tested physical server and the virtual server before and after optimization

before and after the physical test server The TPS optimization is slightly higher than that of virtual servers. However, when User Load ≥ 1000 , the TPS of the physical test server and virtual server before optimization begins to be established quickly, and the average response time is also significantly increased; at present, the TPS of the optimized virtual server also maintains stable development, and there is no average response time. Obviously, the maximum user load of the optimized virtual server is 28% higher than the default value in the other two cases. The TPS is 31% higher than the previous optimization and 25% higher than the tested physical server, as shown in Table 1.

Table 1. Test physical server and specific data of virtual server before and after optimization

User load	Physical server	Optimized virtual server	Virtual server before optimization
100	102	99	101
500	512	446	487
1000	952	832	1024
1500	1485	1099	1521
2000	2014	1524	2201

5 Conclusions

As an emerging information technology, big data technology will lead the future global development and social progress, and promote the progress of human life. This article fully recognizes its importance and uses big data technology to research laboratory information collection systems. Build an IoT laboratory information collection system through system requirements analysis. And test it to become a highly representative and in-depth information management system for colleges and universities. It is planned to rely on big data technology to generate informatized real-time equipment information and create a brand-new Internet of Things laboratory information system.

Acknowledgment. Subject of experimental teaching reform and construction of Huaqiao University in 2021 (SY2021J17).

References

1. Kang, N., Zhao, C., Li, J., et al.: A Hierarchical structure of key performance indicators for operation management and continuous improvement in production systems. *Int. J. Prod. Res.* **54**(21–22), 6333–6350 (2018)
2. Sboui, L., Ghazzai, H., Rezki, Z., et al.: On green cognitive radio cellular networks: dynamic spectrum and operation management. *IEEE Access* **4**(3/4), 1 (2017)
3. Zhang, Q., Yang, L., Ma, C., et al.: Emission characteristics and chemical composition of particulate matter emitted by typical non-road construction machinery. *Atmos. Pollut. Res.* **11**(4), 679–685 (2020)

4. Cai, H., Xu, B., Jiang, L., et al.: IoT-based big data storage systems in cloud computing: perspectives and challenges. *IEEE Internet Things J.* **4**(1), 75–87 (2017)
5. Fajjari, I., Tobagi, F., Takahashi, Y.: Cloud edge computing in the IoT. *Annals of Telecommunications - annales des télécommunications* **73**(7–8), 413–414 (2018)
6. El-Sayed, H., Sankar, S., Prasad, M., et al.: Edge of things: the big picture on the integration of edge, IoT and the cloud in a distributed computing environment. *IEEE Access* **99**, 1 (2018)
7. Fu, J.S., Yun, L., Chao, H.C., et al.: Secure data storage and searching for industrial IoT by integrating fog computing and cloud computing. *IEEE Trans. Indus. Inf.* **14**(10), 4519–4528 (2018)
8. Shah-Mansouri, H., Wong, V.: Hierarchical fog-cloud computing for IoT systems: a computation offloading game. *IEEE Internet of Things J.* **99**, 1 (2017)
9. Bo, W.A., Wei, P.A., He, L.A., et al.: Research on information intelligent collection model of service consultation system - sciencedirect. *Procedia CIRP* **83**, 779–784 (2019)
10. Mo, N., Xie, X., Liang, H.: Design on the farmland planning area information collection system based on cloud computing. *J. Agricult. Mech. Res.* (2020)
11. Xie, X.: Research on the operation management of hema. *World Scient. Res. J.* **5**(7), 97–101 (2019)

Machine Learning, Computer Vision, Image Processing, and Their Intelligence Applications



Three-Dimensional Tree Visualization of Computer Image Data Based on Louvain Algorithm

Chao Mao^(✉)

Division of Science and Technology, Beijing Normal University - Hong Kong Baptist University
United International College, Zhuhai 519000, Guangdong, China

Abstract. At this stage, due to the rise of artificial intelligence, computer vision as an important branch of artificial intelligence is also very popular, and three-dimensional tree visualization is one of the more popular research directions in the field of computer vision. The three-dimensional tree visualization mainly studies how to visualize the process of obtaining three-dimensional information based on a single view or multiple views. The purpose of this article is to study the computer image data three-dimensional tree visualization technology based on Louvain algorithm. In this paper, in the research of 3D tree visualization based on deformable model, the 3D tree visualization research is carried out by using a picture, based on the 3D average model of the known object, and capture the 2D image of the corresponding model and its specific characteristics. The coordinates of the point, iterate continuously until convergence to complete the study. This paper proposes a Louvain algorithm for image data processing. Based on this algorithm, a three-dimensional visualization reconstruction system of atlas is established, and the system structure and implementation process are described in detail. Finally, experiments are carried out to test the system based on the new algorithm, and the data obtained by analysis proves the effectiveness of the Louvain algorithm. Experimental research shows that, combining rendering processing time and data processing time, the data time of the Louvain algorithm used in this paper is 65% shorter than that of the MC algorithm on average.

Keywords: Three-bit reconstruction · Louvain algorithm · Computer image · Three-dimensional tree

1 Introduction

In the past two years, due to the unusual popularity of the Louvain algorithm, there has also been a method of using the Louvain algorithm to visualize a three-dimensional tree of image data on a single picture [1, 2]. Image-based 3D reconstruction technology is an important research direction of computer vision, involving computational geometry, computer graphics, pattern recognition, image processing and many other fields. It has very high research value. It can not only promote the development of related disciplines, but also Promote the rapid development of related applications in the industry.

In the research on the three-dimensional tree visualization of computer image data based on the Louvain algorithm, many scholars have studied it and achieved good results. For example, Strozzi F adopts the method of affine decomposition under the assumption that the camera model is an orthogonal projection. Calculate the three-dimensional structure of the object and other information [3]. Boone J M created a semi-automatic scene 3D reconstruction system through system guidance and manual interaction [4]. It can be seen that it is of great significance to study the three-dimensional tree visualization of computer image data based on the Louvain algorithm.

This paper implements a three-dimensional visualization model of image data based on Louvain algorithm, and runs it on a cluster parallel system built in the laboratory LAN. The algorithm is tested and analyzed from the cluster environment and network speed. In this paper, the three-dimensional tree visualization of image data based on Louvain algorithm. Express the network data by constructing the mapping relationship between the network community and the three-dimensional trees.

2 Research on the Three-Dimensional Tree Visualization of Computer Image Data Based on Louvain Algorithm

2.1 3D Flow Field Volume Rendering of Image Data Type

(1) Conversion of 3D flow field to image data type

The organization form of the three-dimensional flow field data is a structured grid, which is a data organization form with a regular topological structure and an irregular geometric structure. When the data detection (Probe) method is used to convert the organization structure of the flow field data into the image data organization structure, because the data organization structure of the data detection output result is the same as the detection data, and the attribute data is obtained by the attribute interpolation of the detected data. Therefore, this article uses image data as the detection data and three-dimensional flow field data as the detected data. Before data detection, the dimensions, starting coordinates, coordinate point interval and other attributes of the image data must be defined [5, 6]. In order to ensure the applicability of the written program to different flow field data files, when defining the image data attributes, the self-definition of the detection data attributes is completed by extracting the information of the three-dimensional flow field data set.

The density of detection points during data detection needs to be determined according to the actual data density. It is necessary to avoid data redundancy caused by oversampling and also avoid high-frequency information loss caused by undersampling. Therefore, whether the definition of detection data attributes is appropriate or not directly affects The final volume rendering effect. In order to reduce the irrationality of detection density as much as possible based on the realization of image data customization, the dimension of image data, that is, the number of detection points in each coordinate axis direction, is set to the dimension of the three-dimensional flow field [7, 8]. Also, because the three-dimensional arrangement direction of the points in the image data set is consistent with the physical space coordinate system, the spatial extent of the image data set is set to the spatial extent of the flow field. The specific implementation process

of image dataset attribute setting is to first extract the three-dimensional flow field space coordinate range bounds and dimension extent from the flow field data set, and then set the data set attributes through the corresponding calculation formula according to the extracted flow field information.

(2) Texture mapping volume rendering model construction

1) 2D texture mapping model

The rendering efficiency of the 2D texture mapping volume rendering algorithm is related to the software sampling rate, texture loading rate, and texture mapping polygon scanning conversion rate. The texture image needs to be generated during slice sampling. When the processed volume data is stored in a linear sequence, the sampling rate is related to the sampling direction, because this involves the location of the fast buffer [9, 10]. In the process of implementing 2D texture mapping, three sets of slice images can be pre-calculated and sampled along the three direction axes and stored in the hardware buffer, thereby reducing sampling time consumption but increasing memory usage. The texture loading rate refers to the transfer rate from the main memory to the texture mapping memory. The scan conversion of polygons is often limited by the rate at which the graphics hardware processes the pixels in the image.

2) 3D texture mapping model

In theory, the 3D texture mapping volume rendering renderer and the ray casting volume rendering renderer have the same calculation process and complexity, and generate the same pictures, and both methods use the nearest neighbor interpolation algorithm or the cubic linear interpolation algorithm to perform the volume data. For sampling, the same sampling point synthesis method is used to obtain the color value of the pixel. Therefore, the two methods of 3D texture mapping and ray casting can be considered the same in mechanism [11, 12]. The advantage of texture mapping is that it can perform sampling and fusion operations through graphics hardware accelerators, so the speed is faster. However, the use of graphics hardware for volume rendering also has some disadvantages. For example, when the results of each pixel are stored in the sampling point fusion stage, texture mapping will produce more artifacts due to the precision limitation of the frame buffer.

2.2 The Design of the Image 3D Visualization System Based on Louvain Algorithm

(1) Functional requirement design

- 1) In order to facilitate the needs of scientific research, the system must support the reading and access of computer data. In addition, the result data information in the experiment can be saved to the local hard disk, that is, it supports the reading and access of information such as transfer function and gradient.
- 2) The system can perform 3D reconstruction through direct volume rendering algorithm. The system is a comprehensive visualization platform that integrates many classic direct volume rendering algorithms, such as ray casting algorithms, MIP, MIDA, feature enhancement, etc.; at the same time, lighting can be added to enhance

the rendering results in the rendering process. Before drawing the image, necessary preprocessing of the volume data can enhance the stereo perception of the drawing result.

- 3) For the rendered three-dimensional image, the system can support the user's interactive operation, can rotate, zoom and move the image, and support mouse operation of the image, which is convenient and intuitive to observe the organizational details in the image. In order to facilitate comparison, the system can support screenshot operations of rendering results and save the corresponding volume rendering results. In addition, the transfer function is also very important to the drawing result. Users can adjust the transfer function, and the adjusted parameters can be updated to the drawing result in time. It also supports the adjustment of the sampling step size and the setting of the lighting parameters.
 - 4) In order to make up for the limitations of data processing on many platforms, this system supports the visualization of brain white matter fiber data. For this kind of data, the system can add anti-aliasing and other effects in the drawing process to improve the drawing results. In order to reflect the white matter fiber trajectory more realistically, the system can set relevant parameters to effectively eliminate the wrongly tracked fibers and draw high-confidence fibers.
- (2) Layer interface
- 1) Bottom management interface: Located between the bottom layer and the operation layer, it is responsible for the data transmission function between the two. After the bottom layer of the system obtains the original volume data, it provides data to the upper layer through this interface, so that the bottom layer can provide a unified data format to the outside, avoiding drawing failures due to confusion in the data format. At the same time, the bottom layer of the system also provides a set of 3D rendering algorithms upwards through this interface for use by the operation layer.
 - 2) Image rendering interface: After obtaining standardized image data, a certain rendering method needs to be used to perform three-dimensional rendering of the data. The image rendering interface defines two sub-processing interfaces, data preprocessing and image synthesis. These two functional modules are related. The data provided by the bottom layer must first undergo data preprocessing such as denoising and smoothing to wash the volume data. Then, according to the volume rendering algorithm provided by the lower layer, the image synthesis is performed according to certain rules to realize the communication between the bottom layer of the system and the operation layer. At the same time, this interface design is convenient to improve the algorithm of a certain module and add new algorithms, hide the specific implementation details of the volume data processing process, reflect the object-oriented design principle, and facilitate the modular integration and management of the system.
 - 3) Parameter management interface: After the image is drawn, the system provides parameter settings such as sampling step size, illumination coefficient, etc. The parameter management interface is above the operating layer, and the corresponding parameter interface is provided upwards for its invocation. At the same time, it also accepts the parameter results transmitted from the interface layer and performs storage operations. At the same time, the interface also needs to process the interactive results of the transfer function on the display layer. This interface can uniformly

manage the parameters adjusted and processed by this layer and upper-layer users, and realize the preservation and recall of important temporary parameters.

- 4) The top-level interface of the system: At the interface display layer, users need to perform necessary interactions based on the drawn three-dimensional images. The top-level interface provides an interactive interface to the interface, which is convenient for users to use. Using this interface, users can independently define color transfer functions, modify color schemes for volume data, set different opacity for different organizational structures, and highlight organizational details of interest. In addition, the interface also provides operations such as image movement, flipping, etc., to meet basic image processing requirements.
 - a. Module design
 - 1) Volume data processing. For the medical volume data obtained by electronic computed tomography (CT) or nuclear magnetic resonance (MRI), it needs to be denoised pre-processed before the experiment, and converted into a three-dimensional texture and stored in the memory. In order to classify the volume data, it is also necessary to import the adjusted transfer function into the video memory for use;
 - 2) The drawing of the bounding box. In order to obtain the spatial range of volume data, we adopt a bounding box strategy to determine. First, draw the front and back surfaces of the volume data bounding box, so as to obtain the starting point and ending point coordinates of the projected light entering and exiting the volume data. Then subtract the coordinates of the start point from the coordinates of the end point to get the direction and length of the projected light;
 - 3) Image drawing. Connect the pre-written vertex shader and fragment shader, use the advantages of graphics hardware to perform three-line interpolation on the volume data to obtain the attribute value of the sampling point, and obtain the corresponding color and opacity by querying the transmission function; Accumulate light from the sampling points to draw the final image and display it on the screen.

3 Three-Dimensional Tree Visualization of Computer Image Data Based on Louvain Algorithm

The visual parameters of branches include branch angle, length, radius, etc. Among them, branch angle has the greatest impact on the three-dimensional spatial pattern of trees. The branch length r is determined according to the number of nodes in the corresponding community. The more the number, the longer the branch; the azimuth φ is calculated by the Fibonacci sequence. The formula is as follows:

$$\mu_i = \frac{q_i \times C_i(n)}{\sum_{i=0}^K q_i \times C_i(n)} \times \varphi \tag{1}$$

In the formula, μ_i represents the branch angle of the branch corresponding to the community leader; $q_i \times C_i(n)$ represents the weighted node capacity of i ; for an undirected graph containing N nodes, the formula is as follows:

$$C_i(n) = \sum_{j=1}^N x_{ij} \quad i(\neq j) \tag{2}$$

$C_i(n)$ represents the degree centrality of node i , and $\sum_{j=1}^N x_{ij}$ is used to calculate the number of direct connections between node i and other $N-1$ j nodes ($i \neq j$, excluding the connection between i and itself).

4 Experimental Research on Computer Image Data Three-Dimensional Tree Visualization Based on Louvain Algorithm

4.1 System Efficiency Analysis

The operating system of this system is Win10 64-bit platform to process the experimental data. Through the analysis of the efficiency data, it is found that the Louvain algorithm has higher data processing efficiency than other algorithms in processing test data, and the rendering efficiency is the same, which greatly reduces the total time of 3D rendering.

4.2 Test Method

This article mainly discusses how the new Louvain algorithm improves the performance of the entire system, so the total time only involves these two parts. The data of the number of image generation grids is used to judge the consistency of the three-dimensional images generated by the Louvain algorithm and the MC algorithm. The images generated by the system are difficult to distinguish with the naked eye. The three-dimensional image is actually composed by stitching countless small triangular faces, and these small faces are triangular grids before filling. By comparing the number of triangular grids in the three-dimensional image, the two algorithms can be compared to a certain extent. The degree of similarity of three-dimensional images can be obtained by using this method to get the difference in image accuracy between the original MC algorithm and the Louvain algorithm.

5 Experimental Research and Analysis of Computer Image Data Three-Dimensional Tree Visualization Based on Louvain Algorithm

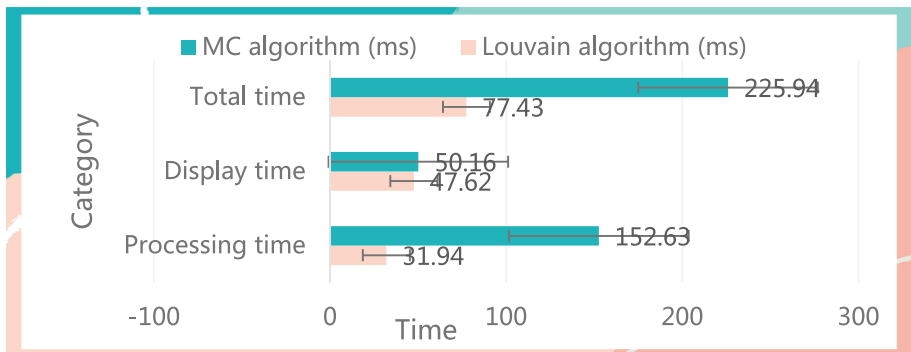
5.1 Comparison of Algorithm Rendering Efficiency

Table 1 shows the overall efficiency comparison between Louvain algorithm and MC algorithm. In terms of rendering processing efficiency, the time consumption of Louvain algorithm is slightly lower than that of MC algorithm. The rendering processing time of each image area is basically stable, and the rendering time of the three-dimensional graphics display of the brain area with the same scale is stable at 50–210 ms.

Table 1. Efficiency comparison between Louvain algorithm and MC algorithm

Category	Louvain algorithm (ms)	MC algorithm (ms)
Processing time	31.94	152.63
Display time	47.62	50.16
Total time	77.43	225.94

As shown in Fig. 1, it can be seen that the Louvain algorithm has basically no improvement in the efficiency of rendering processing, which is also consistent with the previous prediction, because the Louvain algorithm only reduces the number of voxels traversed by the algorithm and basically has no effect on the final generated 3D geometric data., So the display efficiency is not improved. Combining rendering processing time and data processing time, the data time of Louvain algorithm is 65% shorter than that of MC algorithm on average.

**Fig. 1.** The efficiency comparison of Louvain algorithm and MC algorithm

5.2 Three-Dimensional Reconstruction Analysis of Image Data

In this paper, based on the key frame screening of the reconstruction data set, multiple sets of 3D reconstruction experiments have been done. The following mainly shows the results of the four sets of data sets, including the school data set, the globe data set, the wine bottle data set and The student's face data set, the last three data sets were made by shooting with ordinary cameras. The specific content is shown in Table 2:

Table 2. Experimental results time of the algorithm

Group	Louvain algorithm	MC algorithm
School	319	492
Globe	331	451
Vase	94	164
Human face	475	499

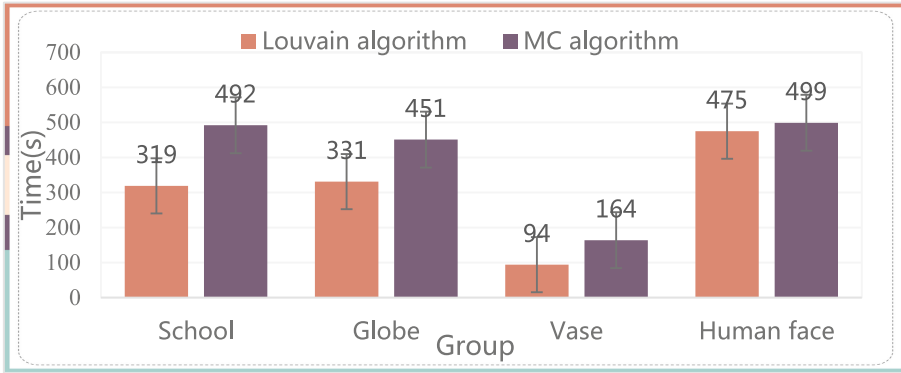


Fig. 2. The experimental results time of the algorithm

As shown in Fig. 2, according to the reconstruction results, it can be known that the algorithm provided in this article can improve the reconstruction efficiency while maintaining the reconstruction results, but the effect is not ideal during the reconstruction of individual special data such as face data. Because the human face is flexible and variable and has some characteristics such as light reflection, there will be many reconstruction obstacles in the process of using traditional motion-based restoration structures.

6 Conclusions

This paper mainly proposes a Louvain algorithm for map data processing. Based on this algorithm, a three-dimensional visualization reconstruction system of atlas is established, and the system structure and implementation process are described in detail. Finally, experiments are carried out to test the system based on the new algorithm, and the data obtained by analysis proves the effectiveness of the Louvain algorithm.

References

1. Seifikar, M, Farzi, S, Barati, M., Blondel, C.: An Efficient louvain-based dynamic community detection algorithm. In: IEEE Transactions on Computational Social Systems, vol. 99, pp. 1–11 (2020)

2. Xiaolan Hou, W., Muqing, L.B., Yifeng, L.: Multi-controller deployment algorithm in hierarchical architecture for SDWAN. *IEEE Access* **7**, 65839–65851 (2019). <https://doi.org/10.1109/ACCESS.2019.2917027>
3. Strozzi, F., Garagiola, E., Trucco, P.: Analysing the attractiveness, availability and accessibility of healthcare providers via social network analysis (SNA). *Decis. Support Syst.* **120**(MAY), 25–37 (2019)
4. Boone, J.M., Hernandez, A.M., Anthony Seibert, J.: Two-dimensional breast dosimetry improved using three-dimensional breast image data. *Radiol. Phys. Technol.* **10**(2), 129–141 (2017). <https://doi.org/10.1007/s12194-017-0404-7>
5. Zeng, W., Chen, G., Rui, J., Yin, H., Tian, W., Tang, W.: The combined application of database and three-dimensional image registration technology in the restoration of total nose defect. *J. Craniofacial Surg.* **29**(5), e484–e487 (2018). <https://doi.org/10.1097/SCS.00000000000004500>
6. Cho, M., Lee, H., Choi, H.H., et al.: A Three-dimensional image transmission using in-network computation in wireless multi-camera networks. *Electron. Dev. Soc. IEEE* **5**(6), 445–452 (2017)
7. Song, Z., et al.: Three-dimensional visualization and measurement of myopectineal orifice in non-inguinal hernia patients. *Surg. Radiol. Anat.* **42**(11), 1315–1322 (2020). <https://doi.org/10.1007/s00276-020-02543-2>
8. Kwon, K.-C., et al.: Advanced three-dimensional visualization system for an integral imaging microscope using a fully convolutional depth estimation network. *IEEE Photon. J.* **12**(4), 1–14 (2020). <https://doi.org/10.1109/JPHOT.2020.3010319>
9. Liu, D., Peng, J., Wang, Y., et al.: Implementation of interactive three-dimensional visualization of air pollutants using WebGL. *Environ. Modell. Softw.* **114**(APR.), 188–194 (2019)
10. Cheng, R., Chen, J., Cao, M.: A virtual globe-based three-dimensional dynamic visualization method for gas diffusion. *Environ. Modell. Softw.* **111**(JAN.), 13–23 (2018)
11. Gargees, R., et al.: Incident-supporting visual cloud computing utilizing software-defined networking. *IEEE Trans. Circ. Syst. Video Technol.* **27**(1), 182–197 (2017). <https://doi.org/10.1109/TCSVT.2016.2564898>
12. Terada, K., Kameda, T., Kageyama, I., Sakamoto, M.: Estimation of three-dimensional long axes of the maxillary and mandibular first molars with regression analysis. *Anat. Sci. Int.* **95**(1), 126–133 (2019). <https://doi.org/10.1007/s12565-019-00506-1>



Sales Forecast for New Energy Vehicles Based on Grey Relevance Analysis and a BP Neural Network

Jing Jiao^{1,2}(✉)

¹ School of Economics and Management, Qinghai Minzu University, Xining, Qinghai, China

² Philippine Christian University Center for International Education, Manila, Philippines

Abstract. New energy vehicles are not only regarded as a breakthrough in the transformation and escalation of China's automobile industry, but also as the only effective way for China to transfer from big automobile country to a strong automobile country. As such, the development of the new energy vehicles industry is of significant importance. In this context, new energy vehicles sales are considered to be one of the key indicators for measuring the new energy vehicles industry development. This paper adopts the actual data on annual new energy vehicles sales from 2011 to 2020. From this data set, seven factors influencing new energy vehicles sales are extracted through literature analysis. Notably, the grey correlation degrees of the selected seven influencing factors are all above 70%. An annual new energy vehicles sales model is then built according to a BP neural network using the seven influencing factors as input variables. The results indicate that the MAPE of the BP model is only 3.15%, meaning that the BP neural network has high accuracy when predicting new energy vehicles sales. By using a BP neural network, it is predicted that new energy vehicles sales volume in China will increase year by year over the next three years, with annual new energy vehicles sales reaching 3.5 million by 2023.

Keywords: New energy vehicles sales · Grey relevance analysis · BP neural network

1 Introduction

Due to the quick development of automobiles, China's automobile industry scale has been ranked first in the world for many years. However, various toxic emissions of automobiles caused greatly dangerous consequences for the environment [1]. Problems such as energy lack, climate change and environmental pollution driven by the greenhouse effect have continued to intensify. New energy vehicles are becoming one of the most effective ways to greatly alleviate this series of problems, which are also essential for China to meet international climate change obligations greenhouse gas emission reduction targets [2]. New energy vehicles are not only regarded as a breakthrough in the transformation and escalation of automobile industry in China, but also as the only way for China to transfer from a big automobile country to a strong automobile country. With

this in mind, it is very essential to do new energy vehicles research, and the sales of new energy vehicles are an important indicator reflecting the development of new energy vehicles. On this basis, this paper aims to analyze and forecast the new energy vehicles volume in China.

In order to analyze new energy vehicles sales, it is necessary to first identify the influencing factors. Wang ZH found that the most important factors influencing Chinese consumers' purchase decisions include the intrinsic product quality, government policy incentives, costs, reference groups and symbolic factors [3]. Guan X investigated whether increasing government subsidies, reducing the life cycle cost and improving the effectiveness would increase consumers' willingness to buy new energy vehicles [4]. Meanwhile, CINZIA C analyzed that vehicles purchase price, charging range, fuel economy and price to determine the extent to which they affect consumers' purchase decisions relating to new energy vehicles by adopting a preference survey [5]. Wang ZY found that seasonal factors, changes in oil prices, development level of new energy technology and policy influence significantly affect consumers' new energy vehicles purchase decisions [6]. On a similar note, Mabaire AM pointed out that the key important factors affecting purchase behaviour of environment-friendly automobiles include environmental knowledge, environmental values and the responsibility feeling [7]. The establishment of a model of new energy vehicles is now mainly divided into two: methods based on mathematical statistics and methods based on machine learning. Mathematical-statistical methods such as Multiple Regression Analysis and the Holt-Winters method rely on historical product sales data [8, 9]. However, it is difficult to grasp market demand and product life cycles through these methods [8], such that their prediction accuracy is not high. The prediction effect of machine learning methods, such as BP neural networks, Bass model, and Fruit Fly optimization algorithms, is superior to mathematical statistical methods [6], and existing research shows that BP neural networks have higher accuracy in predicting sales volume [6, 10]. In recent years, many scholars have combined different methods to solve the large forecast errors caused by a single technology and method, the results of which show an increased forecast accuracy relative to when a single technology is used.

Therefore, in order to obtain higher forecast accuracy, this article combines grey correlation and neural network analysis to forecast the new energy vehicles sales. After extracting and verifying the influencing ability of the key new energy vehicles sales through grey correlation analysis, the influencing factors are then used as input variables to build a BP model to provide the model with high prediction accuracy.

2 Introduction of Related Theories

This paper reviews existing studies to analyze the factors influencing new energy vehicles sales by using grey relevance analysis and calculating the related degree among the influencing factors and actual new energy vehicles to clarify the effective factors. These factors will then form the basis of a BP neural network to guarantee the high prediction accuracy of the model.

2.1 Grey Relevance Analysis

Grey relevance analysis (GRA) is a method employed to analyze the similarities and differences of development degree among various influencing factors, which has been successfully applied to many fields including engineering, management science to verify the influencing ability of each factor [11].

The related formulae are as follows:

$$x_0 = \{x_0(k) | k = 1, 2, \dots, n\} = (x_0(1), x_0(2), \dots, x_0(n)) \tag{1}$$

$$x_i = \{x_i(k) | k = 1, 2, \dots, n\} = (x_i(1), x_i(2), \dots, x_i(n)) \tag{2}$$

$$\gamma_i(k) = \frac{\min_s \min_t |x_0(t) - x_s(t)| + \eta \max_s \max_t |x_0(t) - x_s(t)|}{|x_0(t) - x_i(t)| + \eta \max_s \max_t |x_0(t) - x_s(t)|} \tag{3}$$

k represents the time, x_0 represents the reference sequence; x_i represents the comparison sequence; $\gamma_i(k)$ represents the grey correlation coefficient of the sequence x_i to x_0 at time k; η represents the resolution coefficient, the value interval It is [0,1]; $\min_s \min_t |x_0(t) - x_s(t)|$ represents a two-level minimum difference; $\max_s \max_t |x_0(t) - x_s(t)|$ represents a two-level maximum difference.

$$r_i = \frac{1}{n} \sum_{k=1}^n \gamma_i(k) \tag{4}$$

The correlation degree of the sequence x_i to the reference sequence x_0 is calculated using formula (4).

2.2 BP Neural Network

A BP neural network is a parallel, complex, nonlinear, dynamic algorithm system proposed by Rinehart and McClelland in 1986 [10], A BP model is a multi-layered feed-forward network making use of the error back propagation, which can represent extremely complex nonlinear model systems. Currently, BP neural networks are one of the most widely used neural network models [6]. They have been used extensively and achieved outstanding results in terms of regression, classification, pattern recognition, function approximation and other aspects [10]. As Fig. 1 shows, BP neural network topology consists of input layers, at least one hidden layer, and output layers. The transfer function of the neural node in the network is a sigmoid function, the output of which is successive data from 0 to 1, which can be implemented from the input layer to the output layer of arbitrary non-linear mapping. Although the layers are fully connected, there is no connection among the nodes in the same layer and each layer can only accept the input of the previous layer, meaning it cannot cross the layer.

The existing research has demonstrated that a neural network with only one hidden layer has the ability to approximate any nonlinear continuous function [6]. On this basis, the present paper chose a classical three-layer BP network topology only with one hidden layer, then built a BP model using Matlab.

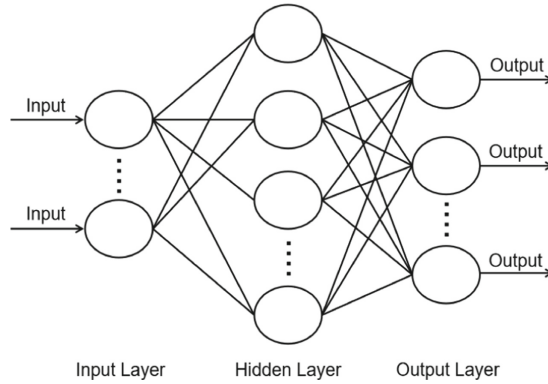


Fig. 1. Three-layer BP neural network structure

3 Model Solving

3.1 Grey Relevance Analysis

Through a literature review and statistical analysis, the following influencing factors are selected: gross domestic product, consumer price index, number of national policies, number of patents in the new energy vehicles industry, number of public charging piles, power battery output, and gasoline prices. Among these, economic level and consumer purchasing power are the fundamental factors that determine the purchasing power of new energy vehicles. Additionally, we use consumer price index and gross domestic product to check economic level. The technical level affects the development and improvement of new energy vehicles, which is measured by the number of patents in the new energy vehicles industry and battery power output. Meanwhile, the number of national policies and public charging piles reflect the state's support for the new energy vehicles. Finally, gasoline prices also affect consumers' willingness to choose to purchase new energy vehicles (Table 1).

Table 1. Grey relevance degree between annual new energy vehicles sales and influencing factors

Influencing factors	Grey relational degree	Rank
Gross domestic product	0.776	4
Consumer price index	0.775	5
Number of patents in the new energy vehicles industry	0.783	2
Power battery output	0.911	1
Number of national policies	0.780	3
Number of public charging piles	0.713	7
Gasoline prices	0.774	6

It can be seen from the table above that the grey correlation degree of the selected seven influencing factors is all above 70%. This indicates that the influencing factors are positively correlated with the sales sample.

3.2 BP Neural Network Analysis

A BP model is built using Matlab R2020a. This network utilizes domestic product, consumer price index, number of national policies, number of patents in the new energy vehicles industry, number of public charging piles, power battery output, and gasoline prices as input variables, whilst new energy vehicles sales is the output variable. The results show that there are 19 hidden nodes, the learning rate is 0.01, there are 5000 iterations, and 1000 loops. The fitting and prediction results are as follows:

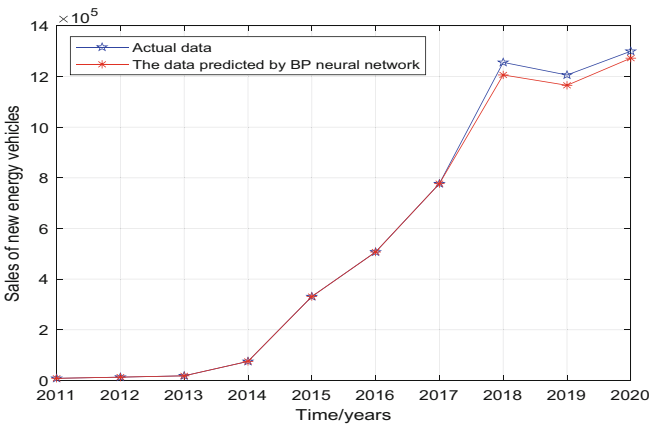


Fig. 2. Fitting and prediction graph of BP model

The blue line means the true value of the sample data, whilst the red line segment is divided into two stages. The first stage means the fitting value of the sample data from 2010 to 2017, whilst the latter stage means the predicted value of the data from 2018 to 2020. Just as Fig. 2 shows, the fitted value is largely consistent with the true value, and there is a very little error between the predicted value and the true value.

Table 2. The error indicator of the BP neural network

Content	MAPE	MAD	MSE	MSE
Score	0.031465	39267.91	1625287072.1174	40314.8493

From Table 2, it shows the error of the BP model is small and acceptable. This is especially true of the MAPE which is only about 3.15%, meaning it is sufficient to meet the prediction requirements. Therefore, the accuracy of the BP model is high.

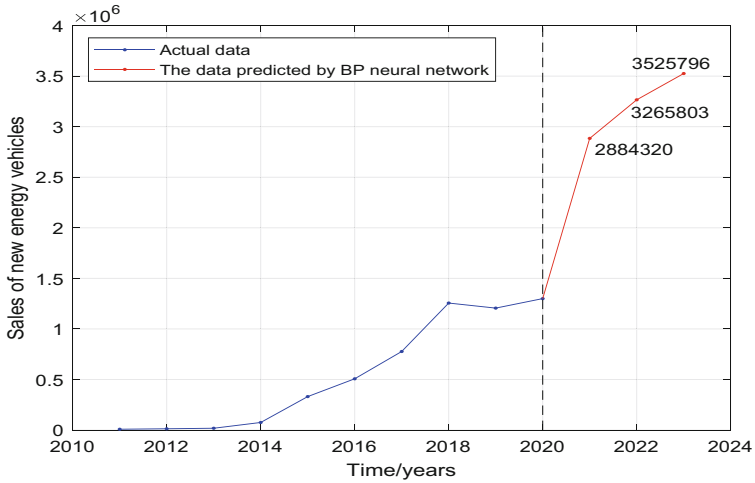


Fig. 3. New energy vehicles sales in China

Finally, the BP model is used to predict the new energy vehicles sales from 2021 to 2023. As Fig. 3 shows, the annual new energy vehicles sales volume in China will increase year by year for the next three years and the new energy vehicles sales will finally attain 3.5 million in 2023, thus indicating that new energy vehicles have broad development prospects in China.

4 Conclusions

This paper calculates the related degree between the seven selected influencing factors and real new energy vehicles sales. It first proves that the seven influencing factors are highly correlated and then establishes BP neural network using the seven influencing factors as input variables. The results show that the trend of prediction results is largely consistent with the actual data and the error is relatively small. Therefore, the BP neural network has both high accuracy in predicting new energy vehicles sales and wide application prospects. It is predicted that the new energy vehicles sales volume will increase year by year for the next three years by using the BP model. Additionally, the new energy vehicles sales volume in China will finally attain 3.5 million in 2023, which means new energy vehicles have huge room for growth. These results have practical reference value for the decision-making of government departments, the production and operation of the automobile industry, and the promotion of new energy vehicles.

References

1. Issakhov, A., Omarova, P.: Modeling and analysis of the effects of barrier height on automobiles emission dispersion. *J. Clean. Prod.* **296**, 1–6 (2021)
2. Kendall, M.: Fuel cell development for New Energy Vehicles (NEVs) and clean air in China. *J. Clean. Prod.* **28**(2), 113–120 (2018)

3. Wang, Z.H., Wang, C., Hao, Y.: Influencing factors of private purchasing intentions of new energy vehicles in China. *J. Renew. Sustain. Energy* **5**(6), 063133 (2013)
4. Guan, X., Zhang, G.X., Liu, D.: The behavior of consumer buying new energy vehicles based on stochastic evolutionary game. *Filomat* **30**(15), 3987–3997 (2016)
5. Cirillo, C., Liu, Y., Maness, M.: A time-dependent stated preference approach to measuring vehicle type preferences and market elasticity of conventional and green vehicles. *Transp. Res. Part A: Policy Pract.* **100**, 294–310 (2017)
6. Wang, Z.Y., Guo, D.H., Wang, H.Y.: Sales forecast of Chinese new energy vehicles based on wavelet and BP neural network. In: *International Symposium on Distributed Computing and Applications for Business Engineering and Science*, no. 10, pp. 141–144 (2019)
7. Mabaire, A.M., Guangquan, X.U.N., Moyo, N.: Purchase behaviour of environment-friendly automobiles. *J. Bus. Soc. Sci. Rev.* **9**(1), 65–72 (2021)
8. Apadula, F., Bassini, A., Elli, A., Scapin, S.: Relationships between meteorological variables and monthly electricity demand. *Appl. Energy* **98**, 346–356 (2012)
9. Tratar, L.F., Strmčnik, E.: The comparison of Holt-Winters method and Multiple regression method: a case study. *Energy* **109**, 266–276 (2016)
10. Kuo, R.J., Tseng, Y.S., Chen, Z.-Y.: Integration of fuzzy neural network and artificial immune system-based back-propagation neural network for sales forecasting using qualitative and quantitative data. *J. Intell. Manuf.* **27**(6), 1191–1207 (2016)
11. Javed, S.A., Khan, A.M., Dong, W., Raza, A., Liu, S.: Systems evaluation through new grey relational analysis approach: an application on thermal conductivity—petrophysical parameters' relationships. *Processes* **7**(6), 348–367 (2019)



Formative Evaluation of College English Autonomous Learning Based on Whale Optimization Algorithm

Xueqian Chi(✉)

School of Foreign Languages, Hulunbuir University, Hulunbuir 021008, China

Abstract. With the advent of the Internet age, online autonomous learning as a new learning method is being accepted by more and more students. With the passage of time, the insufficiency of the network self-learning evaluation mechanism has become more and more obvious, which has affected the creation of teachers' learning resources and the improvement of students' learning effects. This article aims to study the formative evaluation of college English autonomous learning based on whale optimization algorithm. Based on the analysis of the whale optimization algorithm, the characteristics of network autonomous learning, the functions of formative evaluation and the tools of formative evaluation, it is necessary to understand formative evaluation. In the application effect of college English network autonomous learning, this paper selects students from a class of a certain university as the research object, and conducts three-month network autonomous learning in this class. The experimental results show that formative assessment can improve students' autonomous learning ability, and it is also helpful to apply the learned knowledge in practice.

Keywords: Whale optimization algorithm · College English · Internet autonomous learning · Formative assessment

1 Introduction

Evaluation is an important part of education management. Evaluation can guarantee the level of education, improve the quality of education, and is of great significance to the realization of educational goals [1, 2]. Learning self-assessment also needs to adapt to the rapid development of the Internet and the popularization of information technology to bring about major changes in learning tools, methods, environment and other human factors, forming a web-based self-learning evaluation method [3, 4].

Different scholars at home and abroad have different views on formative evaluation. Some scholars indicate that if evaluation is an important part of the introduction of education, it can be regarded as formative evaluation. Formative assessment helps students understand their learning goals, provide feedback, and discover and solve problems. This enables students to reflect on the learning process and evaluate themselves [5, 6]. Some researchers interpret formative evaluation as a process by which teachers recognize and respond to students' learning process during the learning process and improve students'

learning ability [7, 8]. Some researchers interpret formative evaluation as a diagnostic evaluation method that can provide teachers and students with feedback on the educational process. The timing of formative assessment is very random because it provides direct information about student learning [9, 10]. Some researchers have explained the definition of formative evaluation more comprehensively. That is, every activity carried out by teachers and students can be called formative assessment, because the information it provides can be seen as feedback to change educational activities. Compared with other definitions, this definition is easier to be actually understood and applied by teachers and students [11]. In addition, researchers believe formative assessment is done in the education and learning process. The information collected in the assessment can be used to adjust and better adapt to the needs of students and improve the quality of teaching [12]. In addition, formative evaluation has a special function, that is, the definition includes the concept of feedback. In other words, the information that can be obtained from the evaluation process is very different from the overall evaluation.

Based on consulting a large number of references related to formative assessment and autonomous learning, this article combines the whale optimization algorithm, the characteristics of network autonomous learning, the function of formative assessment and the tools of formative assessment, and selects a class in a university as the research object, students in this class will have a three-month online self-learning in order to understand the application effect of formative assessment in college English self-learning.

2 Formative Evaluation of College English Autonomous Learning Based on Whale Optimization Algorithm

2.1 Whale Optimization Algorithm

Whale optimization algorithm is currently a relatively new intelligent optimization algorithm, which is widely used in engineering applications. An extensive study involved 29 test modes and 6 structural engineering problems to evaluate development, exploration potential, local optimal avoidance and convergence. Experimental results show that the whale optimization algorithm is sufficient and superior to other advanced optimization algorithms.

The main steps of the whale optimization algorithm are as follows:

(1) Surround the prey

Whales can locate and surround their prey. Once the best searcher is determined, other searchers will try to move to the best position. The behavior of enveloping can be described as:

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_t - \vec{X}(t) \right| \tag{1}$$

$$\vec{X}(t + 1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \tag{2}$$

In the formula, t is the current iteration, and X is the current position of the whale.

(2) Bubbling offense

In order to describe the mathematical model of the bubbling behavior of humpback whales, two ways are designed, as shown below:

1) Shrink enveloping mechanism

This behavior can be achieved by reducing \vec{a} , setting \vec{A} to be a random number on $[-1, 1]$, so that a new position can be selected between the initial value and the best position.

2) Spiral update position

In this way, whales and prey use spiral transformations to represent their positions. The following formula simulates this behavior, as shown below:

$$\vec{X}(t+1) = \vec{D} \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (3)$$

Among them, b is used to determine the form of the logarithmic spiral, and l is a random number located between -1 and 1 .

In order to simulate this concurrency, WOA has a 50% chance to choose the shrink winding mechanism and spiral model.

3) Search for prey

Whales use random methods to analyze the location of other whales and look for prey. During the search phase, the search engine location is also used randomly. This randomly selected value is used to replace the best result. It emphasizes search capabilities, so that WOA has better global search capabilities.

2.2 The Characteristics of Network Autonomous Learning

(1) Active learning based on web resources

The popularization of the Internet has surpassed the limitations of distance and learning space, providing students with open educational content and exchanging high-quality educational resources for independent learning. If students want to learn, as long as they can connect to the Internet, they can easily and quickly need the resources, which can provide students with the same learning opportunities and give them the opportunity to achieve “educational equity”. This is the best way to achieve lifelong learning.

(2) Relatively independent personalized interactive learning

Online autonomous learning transforms traditional pedagogical education into learning-based learning. Each learner is a relatively independent learner with his own

learning characteristics. Each learner can use platform interactive resources for personalized learning according to their own learning needs. At the same time, the platform can also reflect students' learning conditions in a timely manner, and allocate reasonable learning resources, thus presenting the education rules that teach students according to their aptitude and vary from person to person.

(3) Virtual autonomous learning environment

The Internet creates a virtual learning environment for autonomous learners. In this environment, students can effectively reproduce the face-to-face scenes of traditional classrooms and teachers, and the multi-directional communication between students and learning-related elements. At the same time, the application of technology can provide users with real-time data feedback and has strong flexibility. The display of multimedia courses and on-site demonstration of materials are also conducive to students' easy and quick access to knowledge content and resources.

(4) Flexible learning of independent choice

The starting point and purpose of autonomous learning is to coordinate the interaction between oneself and various factors in learning activities during the learning process. An independent online learning support system on the Internet can learn independently according to one's own needs, flexible, time-saving and labor-saving to achieve the best learning effect, meet learning needs, and optimize learning results. Generally speaking, the more autonomous the study, the closer the learning process is to the student himself, and the better the final learning effect.

2.3 Functions of Formative Evaluation

- (1) Formative assessment can monitor or guide the learning process of students. This is very important, teachers and students can monitor learning progress and guide education and learning plans.
- (2) Formative evaluation focuses on helping language learners understand their own strengths and weaknesses in learning ability. Once they understand their strengths and weaknesses, language learners can use some useful strategies to improve their learning abilities. Formative assessment can determine whether and how much knowledge a language learner has acquired, can cultivate the confidence and interest of language learners.
- (3) Formative evaluation allows teachers and students to discover problems in time. Formative assessment can show teachers what educational content needs to be relearned, and inform language learners what knowledge they have gained and what knowledge they have lost. Once they understand their learning situation, they can improve their language learning in many ways.

2.4 Tools for Formative Assessment

(1) Classroom observation

Observation is “the basis for assessing human skills and behavior.” Classroom observation is the basic way to directly obtain teaching data in the classroom. Only under constant observation will you receive information about student behavior and learning conditions in the classroom. Through classroom observation, teachers can obtain basic knowledge of student performance, which helps teachers design, implement and modify their educational plans.

(2) Portfolio assessment method

Since the ultimate goal of formative assessment is to improve learning, many tools and techniques are used to support the assessment. The portfolio assessment method is a useful tool to improve the quality of formative assessment. Its purpose is to measure the abilities of students and inform teachers of their strengths and weaknesses. An ideal learning portfolio will hold the key information of students’ learning and development, and paint a dynamic, comprehensive and three-dimensional development picture for students, so that students can feel the progress in their learning life, thereby enhancing their self-confidence in learning. In the process of creating a learning portfolio, students can independently participate in the selection of the content of the learning portfolio. By constantly recalling the contents of the portfolio, self-reflection, continuous improvement, and the pursuit of excellence can be realized, so that students can avoid relying on the guidance of teachers, and indirect passive students gradually become independent students.

(3) Diary

Diary is another formative assessment tool, also called dialogue diary, which records the dialogue between teachers and students. Both teachers and students can write diaries. It is a means for teachers and students to communicate and communicate emotions and ideas to a certain extent. It helps to improve the relationship between teachers and students.

3 Experiment

3.1 Research Objects

This article takes a class of non-English majors in a university as the research object. The class has 48 students. In this class, 5 h of online self-learning time are arranged every week. The experiment process lasts for 3 months. Before the experiment, investigate the students’ English autonomous learning ability. After the experiment, they will be investigated again and compared with the results before the experiment to understand the application effect of formative assessment in college English autonomous learning on the Internet.

3.2 Implementation Process

- (1) To apply formative assessment to online autonomous learning, learners should first know and share their own learning goals and performance standards. English teachers can clarify to all students the teaching goals and characteristics of college English courses in this semester in the first class of each academic year, and define or suggest corresponding online self-learning content. Of course, teachers can also form an incentive mechanism to discuss with students and establish their own learning goals and success criteria through the learning goals and success criteria established with students.
- (2) Online self-learning encourages students to practice collaboratively to overcome time and space constraints. Learners can communicate information and network interaction and mutual evaluation during the teaching process. The formative assessment at the student level is divided into four forms: student self-assessment, peer assessment, peer consultation and self-regulation.

4 Discussion

4.1 Formulation and Implementation of Study Plan

Table 1. Development and implementation of student learning plans

	Before the experiment	After the experiment	P
Set goals according to their own reality	5.38 ± 1.00	7.41 ± 1.16	0.028
Self-made study plan	4.91 ± 0.92	7.14 ± 0.80	0.00
Adjust the plan according to the learning situation	4.25 ± 0.67	5.15 ± 0.88	0.034
Plan study time	4.76 ± 1.00	6.42 ± 0.95	0.027

The development and implementation of student learning plans are shown in Table 1 and Fig. 1. From the statistics, it can be seen that in determining the learning goals according to their own reality, the average value of students after the experiment was 7.41, an increase of 2.03 points compared with before the experiment. In the independent development of English learning plans, the data reached 7.14 after the experiment, indicating that after a period of study, students are more active and proactive in making plans. In the third survey, the scores before and after the experiment were 4.25 and 5.15, respectively, indicating that after three months of study, most students have improved in adjusting their learning plans according to their own learning situation, but they are still at a low level. In terms of arranging study time, most students can arrange English study time reasonably, which is better than before.

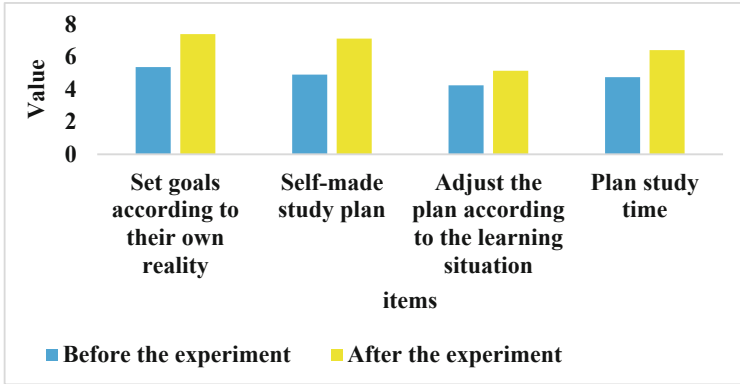


Fig. 1. Development and implementation of student learning plans

4.2 Persistence of Learning

It can be seen from Table 2 and Fig. 2 that in terms of learning persistence, the students were able to persist in English preview and review after the experiment, reaching a good standard; in the second item, the average after the experiment reached 7.51, which was an increase of 2.56 compared to before the experiment. The difference is significant, indicating that after three months of formative evaluation of online self-learning, most students have realized their learning deficiencies and have taken specific actions to remedy them; in insisting on reading every day, the students after the experiment are all worth 9.34, the difference is significant; in applying the knowledge learned in practice, the average values before and after the experiment were 2.78 and 3.33, respectively, indicating that most students in the experimental class still have great deficiencies in applying language knowledge and skills.

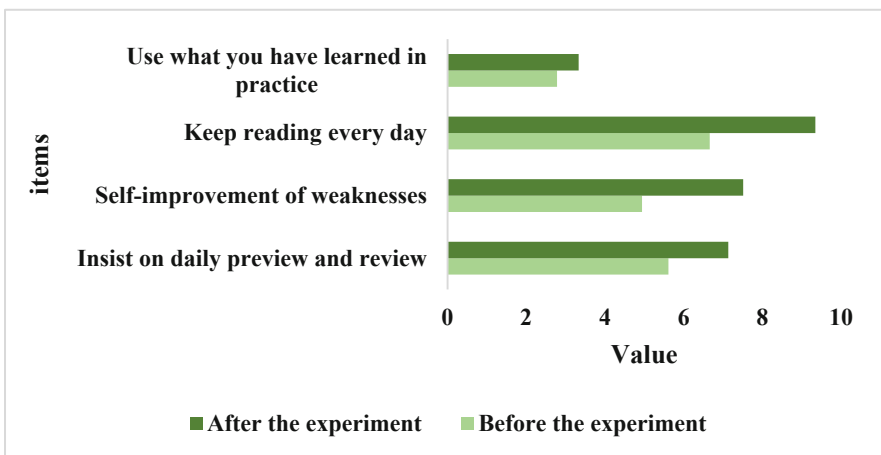


Fig. 2. Analysis of the persistence of students' learning

Table 2. Analysis of the persistence of students' learning

	Before the experiment	After the experiment	P
Insist on daily preview and review	5.61 ± 0.80	7.13 ± 1.66	0.020
Self-improvement of weaknesses	4.94 ± 0.53	7.51 ± 1.64	0.022
Keep reading every day	6.66 ± 1.31	9.34 ± 1.35	0.033
Use what you have learned in practice	2.78 ± 0.20	3.33 ± 0.51	0.086

5 Conclusions

At present, the theoretical research and application practice of autonomous learning is very rich, but it is limited to the autonomous learning in the traditional education environment, while the autonomous learning evaluation research in the new era network environment is relatively small, and the real-time feedback of the evaluation process is insufficient, and the network technology. The degree of integration is not high, and more in-depth theoretical research is needed to better support and practice teaching.

Acknowledgement. Periodic Achievement of the Research Program “Study on Classroom Silent Behavior and Self-regulation of Ethnic Minority Students under the Background of Three-language Acquisition”.

References

1. Baskerville, N.B., Struik, L.L., Dash, D.: Crush the crave: development and formative evaluation of a smartphone app for smoking cessation. *Jmir Mhealth Uhealth* **6**(3), e52 (2018)
2. Schnefke, C.H., Thakwalakwa, C., Muth, M.K., et al.: Optimizing portion-size estimation aids: a formative evaluation in Malawi. *Public Health Nutr.* **22**(17), 1–13 (2019)
3. Cotner, B., Arriola, N., Dirk, L., et al.: The “tools for a working life with spinal cord injury” toolkit: a formative evaluation. *Arch. Phys. Med. Rehabil.* **101**(12), e151–e152 (2020)
4. Muljana, P.S., Luo, T., Watson, S., Euefueno, W.D., Jutzi, K.N.W.: Promoting instructional designers' participation in free, asynchronous professional development: a formative evaluation. *J. Form. Des. Learn.* **4**(2), 74–87 (2020). <https://doi.org/10.1007/s41686-020-00044-4>
5. Yu, Y., Li, W., Li, J., et al.: A novel optimised self-learning method for compressive strength prediction of high performance concrete. *Constr. Build. Mater.* **184**(SEP.30), 229–247 (2018)
6. Kikkawa, N., Kajita, S., Takechi, K.: Self-learning molecular design for high lithium-ion conductive ionic liquids using maze game. *J. Chem. Inf. Model.* **60**(10), 4904–4911 (2020)
7. Chen C, Xiao YX, Liu J, et al. Symmetry Enforced Self-Learning Monte Carlo Method Applied to the Holstein Model. *Phys. Rev. B Condens. Matter Mater. Phys.* **98**(4), 041102.1–041102.6 (2018)
8. Wei, Q., Li, B., Song, R.: Discrete-time stable generalized self-learning optimal control with approximation errors. *IEEE Trans. Neural Netw. Learn. Syst.* **29**(4), 1226–1238 (2018)
9. Rachel, A.R., Sharmila, T., James, G., et al.: Modified directed self-learning sessions in physiology with prereading assignments and Pecha Kucha talks: perceptions of students. *AJP Adv. Physiol. Educ.* **42**(1), 26–31 (2018)

10. Dechanupaprittha, S., Jamroen, C.: Self-learning PSO based optimal EVs charging power control strategy for frequency stabilization considering frequency deviation and impact on EV owner. *Sustain. Energy Grids Netw.* **26**(3), 100463 (2021)
11. Moya, B., Alfaro, I., Gonzalez, D., et al.: Physically sound, self-learning digital twins for sloshing fluids. *PLoS ONE* **15**(6), e0234569 (2020)
12. Yang, Z., Jin, Y., Hao, K.: A bio-inspired self-learning coevolutionary dynamic multiobjective optimization algorithm for internet of things services. *IEEE Trans. Evol. Comput.* **23**(4), 675–688 (2019)



Student Performance Prediction and Learning Intervention System Model Based on Machine Learning

Yuqiang Kong^{1,2}(✉)

¹ Graduate School, Emilio Aguinaldo College, Manila, Philippines

² College of Computer, Guangdong Business and Technology University, Zhaoqing 526060, Guangdong, China

Abstract. Using data mining technology to collect and analyze a large number of learning data generated during the learning process of learners, discover learners' learning characteristics and problems, so that teachers can adopt reasonable intervention measures to learners to improve learners' learning effects. This article aims to study the student performance prediction and learning intervention system model based on machine learning. Based on the analysis of the educational data mining process and performance prediction algorithms, the student performance prediction model and learning are constructed by comparing the performance of four different algorithms. The experimental results show that special attention should be paid to the historical information of students in the study of student behavior and study performance prediction.

Keywords: Machine learning · Data mining · Performance prediction · Learning intervention

1 Introduction

In the context of the information age, all classes of society have accumulated massive amounts of data, and the amount of data has grown exponentially. Valuable knowledge and information are often hidden under massive data, making it a valuable resource for many disciplines and industries [1, 2]. Machine learning and data mining technologies have discovered some unique data laws, and extracted valuable information and knowledge from them to solve problems in different disciplines, allowing managers to make more scientific decisions [3, 4].

Using performance prediction to intervene in student learning has always been a hot research topic abroad, and related research is relatively mature and in-depth. Some scholars used the Apriori algorithm, a classic association rule mining algorithm in data mining, to discover two types of association relationships, namely: the correlation between entrance examination scores and courses. Deriving these two types of correlation can provide useful support for formulating a reasonable teaching plan [5, 6]. Some scholars use correlation algorithms to extract and analyze student performance information,

explore the relationship between courses, and then achieve the purpose of helping college education managers to alert student performance in time [7, 8]. Based on the improvement of decision tree algorithm C4.5, some researchers built a performance prediction model based on the relevant information tree in the student performance database, and realized the prediction of student entrance examination scores [9, 10]. There are also some researchers who analyze the overall performance of students in a specific university to determine the characteristics of student performance distribution, use four evaluation models to evaluate and compare students' learning conditions, and finally decide to use a linear regression model to evaluate and predict student performance [11, 12]. In general, domestic research involves more theoretical research than practical applications, and systems developed by data mining are rarely applied in practice.

On the basis of consulting a large number of related references, combining the process of educational data mining and performance prediction algorithms, this paper constructs a student performance prediction model and a learning intervention model, and compares the performance of the four algorithms in the student performance prediction model. The test results show that the XGBOOST algorithm has the best predictive effect.

2 Student Performance Prediction and Learning Intervention System Model Based on Machine Learning

2.1 Educational Data Mining Process

(1) Data preprocessing

It is difficult for data mining algorithms to directly use these data, because the original data can come from a variety of data sources, and can even include noisy, incomplete or inconsistent data. The data preprocessing is the process of transforming the original data into a new data format that meets the needs of data mining. This involves four major processes: data processing, data integration, data processing specifications and data exchange. Among them, data cleaning technology mainly includes filtering useless and unnecessary data, eliminating noise in big data processing, ensuring data integrity, and filling in missing big data processing through mathematical statistics. The data integration technology is to integrate data from different system databases. The data storage and management protocol is used to increase the effect of big data mining, thereby reducing the size of the original data. The data is processed according to the characteristics of the selected algorithm through data conversion.

(2) Data mining

As a key process in the data mining process, the primary purpose of data mining is to establish a data model. In the data mining stage, it is first necessary to characterize the used scenario problem and the data set itself, and then select the appropriate algorithm to deal with the scenario problem. Therefore, the purpose of the sorting algorithm is to assign types to objects. The purpose of statistical methods and regression algorithms is to assign values to statistical objects. The purpose of the grouping algorithm is to

divide similar data objects in the same category, and the purpose of mining association rules is to find the similarities or associations between data objects. Secondly, if you want to apply the selected algorithm to data mining, you need to continuously refine it throughout the data mining process. After adjusting the parameters, you may have to go back and perform data mining to get the model.

(3) Model evaluation

Model evaluation is the evaluation of the model obtained after data mining. The generated model must be tested in a series of tests, and the quality of the model must be evaluated according to the relevant model indicators. If the indicators do not meet expectations, the model needs to be further optimized until the demand is met.

2.2 Performance Prediction Algorithm

(1) XGBOOST

XGBoost is a massively parallel tool based on tree enhancement. XGBoost reduces computational complexity, speeds up calculations, does not change the input data, has strong interpretability, and is easy to adjust parameters. Whether it is big data or data with a small number of variables, it shows excellent characteristics such as improved accuracy. Solve various supervised learning problems very effectively. Therefore, it is widely admired in various data science competitions, even in the fields of machine learning and data mining.

(2) Gradient boosting decision tree

Gradient Boosting Decision Tree (GBDT) is an improved and optimized algorithm based on the built-in Adaboost learning algorithm. In fact, the promotion method based on the decision tree is called the promotion tree. For classification problems, the decision tree is a binomial classification tree, and for regression problems, the decision tree is a binomial regression tree. The relationship between data input and output is complex. Boosted trees are a powerful machine learning algorithm because the linear combination of decision trees is optimal for training data.

(3) Random forest

Random forest is an improved version of traditional decision tree. The decision tree is constructed retrospectively from top to bottom, so there is no problem with the decision tree creation process. There is no need to replace the data selected from the training data. The process of building a decision tree:

- 1) The input data mainly includes training set features and class labels.
- 2) Choose the most suitable attribute as the root node division attribute.

- 3) For each branch to be split, if it is already in the same category, split it again. If they are not in the same category, select different attributes in turn as the split attributes of the split, and delete the selected split attributes.
 - 4) Repeat step 3) until reaching the leaf node, the last level of the decision tree. At present, all data under this node belong to the same type.
 - 5) Finally, the class label marked by each leaf node is obtained.
- (4) Support vector regression

For traditional regression algorithms, the method of calculating model loss is usually to calculate the difference between the predicted value of the model and the actual value, while support vector regression uses a different idea. In other words, only in the case of absolute model error. Only when the absolute value of the difference is greater than ϵ , the loss of the value and the actual value model is calculated, where ϵ is the maximum deviation that the model can tolerate.

3 Experiment

The purpose of the machine learning task is to obtain a model with strong generalization ability. In order to obtain the predictive ability of the model more objectively and accurately, it is necessary to choose a more appropriate method of evaluating models and indicators. In order to ensure the accuracy of the evaluation of the experimental results, this paper adopts a ten-fold cross-validation method. First, the data set is divided into 10 equal parts. Whenever one of the data sets is used as a test set and the other is used as a training set. Train to receive test results. Repeat 10 times to get 10 test results, and finally take the average of the 10 results as the final result of the model index.

To accurately evaluate the generalization of the model, an index that can fully measure the performance of the model is also needed. This article is a regression task. The commonly used evaluation scores for regression problems are average absolute error, mean square error and coefficient of determination.

The average absolute error is calculated as shown in formula (1), where $f(x^i)$ is the predicted value of sample x^i , and y^i is the true result. It can solve the problem of absolute error offset between positive and negative. The smaller the MAE value, the better the model fitting effect.

$$MAE = \frac{1}{n} \sum_{i=1}^n |f(x^i) - y^i| \tag{1}$$

Mean square error. The calculation method is shown in formula (2). It is used to detect the deviation between the predicted value and the true value.

The worse the test effect.

$$MSE = \frac{1}{n} \sum_{i=1}^n (f(x^i) - y^i)^2 \tag{2}$$

The calculation method of the coefficient of determination is shown in formula (3).

$$R2 = 1 - \frac{\sum_{i=1}^n (f(x^i) - y^i)^2}{\sum_{i=1}^n (\bar{y} - y^i)^2} \tag{3}$$

In the formula, \bar{y} is the mean value of all results. Generally speaking, the larger the R2, the better the model fitting performance.

4 Discussion

4.1 Result Analysis

The features used in the model in this article include diligence-related indicators, behavioral rules-related indicators, the new indicators proposed in this article, and students' past performance. This article conducts ablation experiments and combines quantitative features in different ways to study the model results of different algorithms and Forecast model's demand for different characteristics. This article first creates a predictive model that learns the diligence index, and then continuously adds new features to the model, readjusts the parameters, and evaluates the model. That is equivalent to using the controlled variable method to conduct five different experiments, each experiment using SVR, GBDT, RF, XGBOOST four different regression algorithms to build four prediction models. The experimental results are shown in Table 1:

Table 1. Comparison of ablation experiment indicators

Learner	Index	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
SVR	MAE	0.08299	0.08546	0.08622	0.08587	0.08249
	MSE	0.01501	0.01417	0.01404	0.01384	0.01252
	R2	0.14770	0.17292	0.17547	0.18718	0.26137
GBDT	MAE	0.08154	0.08125	0.08240	0.08100	0.06618
	MSE	0.01464	0.01390	0.01358	0.01339	0.01108
	R2	0.16905	0.19831	0.21825	0.22976	0.37175
RF	MAE	0.08332	0.08271	0.08140	0.08171	0.06482
	MSE	0.01438	0.01338	0.01311	0.01304	0.00950
	R2	0.17284	0.21424	0.23304	0.23315	0.45377
XGBOOST	MAE	0.08337	0.08283	0.08200	0.08118	0.06486
	MSE	0.01418	0.01332	0.01306	0.01281	0.00919
	R2	0.18203	0.22350	0.23534	0.25387	0.47429

It can be seen from Table 1 that for five different experiments, the prediction model indicators constructed by the four algorithms all show relatively consistent results, that is, among the various model indicators, the performance of the SVR indicator is the worst, and the performance of the XGBOOST indicator is the best. This shows that the XGBOOST algorithm in this article has the best prediction results, so this algorithm is selected as the model prediction algorithm in this article.

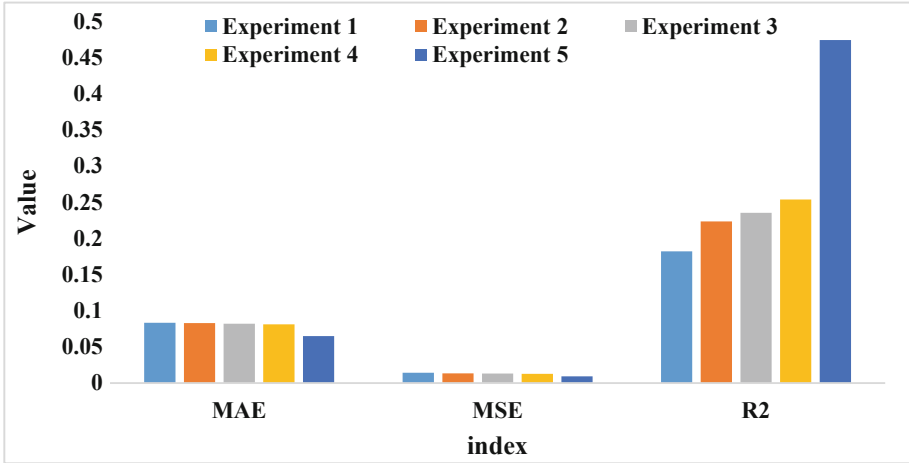


Fig. 1. Comparison chart of indicators of different experiments

This article takes the best XGBOOST model algorithm as an example. We will illustrate the experimental effects of using different feature combinations and draw a comparison chart of indicators as shown in Fig. 1. It can be seen in each round of experiments that the performance of the model has been greatly improved, which shows that the quantified indicators in this paper are of great significance to the construction of accurate prediction models. From Experiment 3 to Experiment 4, we can see that by adding students' past performance information, the performance of the prediction model has been significantly improved. This shows that historical information plays a very important role in constructing predictive models. Therefore, when studying student behavior and predicting academic performance, special attention should be paid to student historical information.

4.2 Learning Intervention Strategies

According to the results of learning performance prediction, an adaptive learning system can be created for learners, and a large-scale learning intervention model in the adaptive learning system, as shown in Fig. 2.

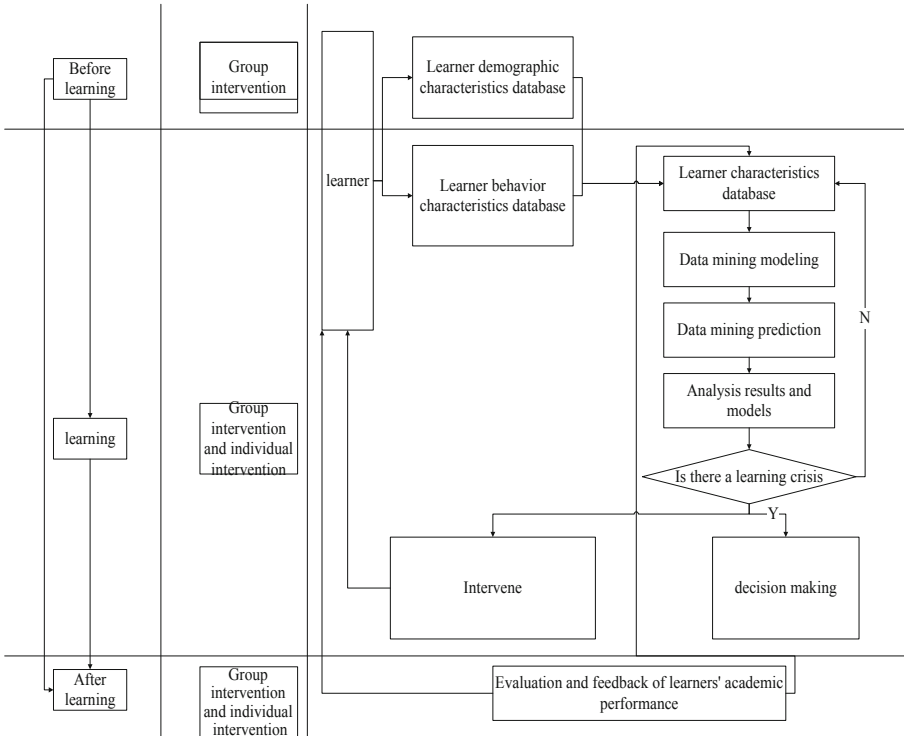


Fig. 2. Learning intervention model of adaptive learning system

(1) Curriculum developer's perspective

Develop adaptive learning systems to improve interventions. The learning progress reminder can clearly show your learning progress and remind you whether the learning meets the standards. At the same time, students can view other students' digital dashboards, allowing them to supervise each other's learning. The Bullet and Reminder windows provide students with real-time or interval reminder feedback and customizable resource upgrades, which are seamlessly connected through SMS, WeChat, and email. Remind students of the performance of learning activities, knowledge content, and behavior characteristics that can be reminded and supervised through multiple channels. Learning Strategy Guidance provides students with learning strategy guidance, including communication and collaboration strategies and self-regulation strategies, especially for

students who are prone to learning disabilities. Behavior participation resource policy driven.

(2) Teacher's perspective

First of all, teachers need to use educational strategies such as group difference guidance to teach according to students' abilities and various reward and punishment mechanisms. The second is to give full play to the leading role of teachers. The advantage of online learning is that you can learn independently, but the disadvantage is that the control ability is weak. Online learning makes students ignorant of time, prone to inaction, delayed interactive behavior, and plagiarism when submitting homework. They will not only affect student behavior, but also affect the effect of online learning.

(3) The learner's perspective

First of all, it is recommended that learners consciously increase their participation in online learning and increase their learning time. Secondly, it is recommended that learners consciously improve their learning strategies, follow the teacher's teaching arrangements, gradually complete learning tasks, clarify learning goals before online learning, formulate a clear learning plan, and complete homework in accordance with the learning plan. They need to participate in discussions, exchanges, review notes and other interactions to gain a sense of success and a sense of belonging to learning and action. They manage, coordinate and evaluate their behavior and psychology in a timely manner.

5 Conclusions

As one of the important research fields of educational data mining, student performance prediction has received widespread attention, and many researchers at home and abroad have successively carried out fruitful research. However, in the existing research work, there is still much room for improvement in the predictability and accuracy of methods for predicting student performance.

References

1. Riihijarvi, J., Mahonen, P.: Machine learning for performance prediction in mobile cellular networks. *IEEE Comput. Intell. Mag.* **13**(1), 51–60 (2018)
2. Dong, S., Zhang, Y., He, Z., et al.: Investigation of support vector machine and back propagation artificial neural network for performance prediction of the organic rankine cycle system. *Energy* **144**(FEB.1), 851–864 (2018)
3. Hao, W., Lu, Y., Lai, Y., et al.: Research on operation strategy and performance prediction of flat plate solar collector with dual-function for drying agricultural products. *Renew. Energy* **127**(NOV), 685–696 (2018)
4. Tirupathinaidu, C., Jain, N., Renganathan, T., et al.: Unified thermodynamic model for performance prediction of adiabatic feedstock gasifiers. *Ind. Eng. Chem. Res.* **59**(44), 19751–19769 (2020)

5. Bao, Y., Lu, F., Wang, Y., et al.: Student performance prediction based on behavior process similarity. *Chin. J. Electron.* **29**(6), 1110–1118 (2020)
6. Al-Faifi, A.M., Song, B., Hassan, M.M., et al.: Performance prediction model for cloud service selection from smart data. *Future Gener. Comput. Syst.* **85**(AUG), 97–106 (2018)
7. Armetti, G., Migliazza, M.R., Ferrari, F., et al.: Geological and mechanical rock mass conditions for TBM performance prediction. The case of “La Maddalena” exploratory tunnel, Chiomonte (Italy). *Tunn. Undergr. Space Technol.* **77**(JUL), 115–126 (2018)
8. Jani, D.B., et al.: A review on use of TRNSYS as simulation tool in performance prediction of desiccant cooling cycle. *J. Therm. Anal. Calorim.* **140**(5), 2011–2031 (2019). <https://doi.org/10.1007/s10973-019-08968-1>
9. Badshah, M., Badshah, S., Jan, S.: Comparison of computational fluid dynamics and fluid structure interaction models for the performance prediction of tidal current turbines. *J. Ocean Eng. Sci.* **5**(2), 164–172 (2020)
10. Peter, W.A., Christian, V., Christoph, R., et al.: The vivarium: maximizing learning with living invertebrates—an out-of-school intervention is more effective than an equivalent lesson at school. *Insects* **9**(1), 1–26 (2018)
11. Sacr, M., Fors, U., Tedre, M., et al.: How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. *PLoS ONE* **13**(3), e0194777 (2018)
12. Schweder, S., Raufelder, D.: Needs satisfaction and motivation among adolescent boys and girls during self-directed learning intervention. *J. Adolesc.* **88**(5), 1–13 (2021)



Construction of Models for Evaluation and Development of Urban Tourism Resources Based on Artificial Intelligence

Chunling Liu^(✉)

Hohhot Vocational College, Hohhot 010051, Inner Mongolia, China

Abstract. With the progress of human society, people pay more and more attention to natural and cultural resources, and urban tourism resources have also received extensive attention. The evaluation and development of urban tourism resources has also become a hot topic in the current research field. Urban tourism has become a new trend in the tourism consumption market, and the tourism industry has huge potential for development and strong growth points. Due to the influence of many factors in the evaluation and development of urban tourism resources, comprehensive analysis of various data is required, which is a complex systematic project. Tourism resources are the material basis and prerequisite for tourism development. Understanding the framework conditions such as the quantity, scope, quality, distribution, and development potential of tourism resources has become a necessary condition for the large-scale development of tourism resources. A reasonable and scientific evaluation of resources is the scientific basis for the formulation of tourism planning and development planning, and ensures the correct formulation of the sequence, depth and direction of the development of tourism resources. At the same time, with the continuous development of science and technology, artificial intelligence has been widely used in various fields. Therefore, by constructing a model for comprehensive evaluation and development of tourism resources to analyze the relationship between various influencing factors, it plays an important role in promoting the sustainable and healthy development of tourism. This article adopts experimental analysis method and data analysis method, and aims to build a systematic and quantifiable model based on artificial intelligence technology to measure the value of urban tourism resources and provide a reference for the development and utilization of its resources. According to the research results, with the technical support of artificial intelligence, the intelligent model of this experiment can basically evaluate the tourism resources of a city, and to a certain extent can reflect the current situation of local tourism and the direction that needs improvement, so as to realize reasonable and effective development to promote the development of the local tourism industry.

Keywords: Artificial intelligence · Urban tourism resources · Resource evaluation · Development research

1 Introduction

With the rapid development of artificial intelligence technology, it is playing an increasingly important role in the national economy, bringing great convenience to people. At the same time, people are paying more and more attention to the development and utilization of tourism resources, and the evaluation and development of urban tourism resources has also become a hot topic in the current tourism industry. The evaluation and development of urban tourism resources is a complex and tedious task. Its purpose is to comprehensively analyze and evaluate various factors such as nature, humanities and society, and propose corresponding measures. Utilizing the advantages of artificial intelligence, high degree of informationization and powerful functions is conducive to constructing an urban tourism resource evaluation and development model, scientifically and rationally evaluating urban tourism resources, and providing a scientific basis for formulating tourism planning and development plans, so as to promote the sustainable and healthy development of urban tourism.

At present, scholars and experts pay more attention to the two aspects of artificial intelligence and tourism resource evaluation and development, and the related research results are also very rich. For example, Chen Yu pointed out that the use of mathematical models to evaluate the city's own tourism resources and the ability to receive tourists can analyze the situation more scientifically and rationally, provide an accurate basis for the development and utilization of tourism resources, and improve economic efficiency [1]. Yu Hu believes that by building a tourism resource database, analyzing the scale, space management, and regional tourism functions and layout, the scientific nature of the development and utilization of tourism resources can be improved [2]. Chen Jianmin proposed to build a smart tourism big data analysis model, which can analyze relevant data efficiently and accurately, and also reduce the cost of tourism service providers, thereby promoting tourism service providers to make improvements in related areas, so as to provide tourists with a comfortable travel experience [3]. Therefore, this article is based on artificial intelligence to conduct research on urban tourism resource evaluation and development model construction, which meets the requirements of the development of the times and has important research value and application significance.

This article mainly discusses these aspects. First, it briefly explained what artificial intelligence is and the background of this era. Then, the evaluation and development of urban tourism resources are elaborated. In addition, the application of comprehensive weighting algorithm in the evaluation and development of tourism resources is also introduced. Finally, based on artificial intelligence, for the construction of urban tourism resource evaluation and development models, experimental research is carried out, and experimental results and analysis conclusions are drawn.

2 Construction of Model of Urban Tourism Resources Evaluation and Development Based on Artificial Intelligence

2.1 Artificial Intelligence

Artificial intelligence is a combination of computer and machine for humans to process information and analyze data to complete the interactive intelligent control process for humans and machines.

With the rapid rise of high-tech disciplines such as computer technology, communication networks, and biological sciences, more and more new concepts of artificial intelligence have emerged, and they have been gradually applied to life and production, and have achieved significant results. At the same time, some difficult-to-achieve or difficult-to-operate work tasks have been better completed. For example, in the field of science and technology, intelligent robots have become an indispensable and irreplaceable tool in the production and life of modern society [4, 5].

Artificial intelligence technology is a complex and highly comprehensive subject, which has applications in industry, service industry, autonomous driving, medical and industry, financial industry, education industry, legal industry, military and so on. In practical applications, artificial intelligence often requires a large number of algorithms for processing. Therefore, a fast, efficient, safe, reliable and stable program is needed to solve complex and high-complexity problems.

The research of artificial intelligence has developed to a certain extent. People can use intelligent robots to complete various complex, tedious and difficult tasks, such as sweeping robots, industrial production robots, etc. With the continuous improvement of computer technology, some domestic enterprises have begun to study artificial intelligence systems to improve their competitiveness and meet market needs.

2.2 Evaluation and Development of Urban Tourism Resources

Tourism resources refer to various things and elements that can attract tourists to nature and human society, can be developed and used for tourism, and can bring economic, social and environmental benefits. The evaluation of tourism resources is based on the principles of science, objectivity, and comprehensiveness. On the basis of seeking truth from facts, one or more scientific and effective methods are adopted to scientifically and reasonably evaluate local tourism resources in terms of the quality of local tourism resources, resource environment and development conditions. The content of tourism resource evaluation is shown in Fig. 1.

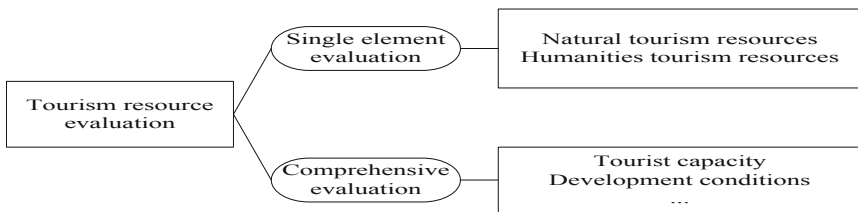


Fig. 1. Contents of tourism resource evaluation

Initially, the evaluation of tourism resources was based on the needs of practical work such as environmental protection and the construction of national parks. With the advancement of science and technology, advanced technologies such as the Internet use analysis and information processing, and spatial positioning and analysis skills have also been widely used in tourism resource evaluation [6, 7].

Through the evaluation of tourism resources, the nature and scale of local tourism resources are determined, and then the key points of developing local tourism resources are pointed out, which has a direct impact on the scientific formulation of tourism planning. First, the scientific evaluation of local tourism resources can provide a theoretical basis for the targeted positioning and ecological planning of the development of local tourism resources. Second, tourism resource assessment can also determine the scientific development direction for the development of local tourist attractions, determine the development direction of tourist attractions, and provide reasonable improvement measures for the developed and partially developed tourist attractions. Third, in terms of management and protection, scientific and reasonable evaluation of tourism resources can make corresponding grade judgments on corresponding tourism resources, thereby determining the level of management and protection, and implementing more effective tourism resource management and protection.

The scientific evaluation of tourism resources must follow the following principles. The first is scientific principles. The research on the value of tourism resources involves multiple disciplines, including geography, aesthetics, sociology, economics, etc. Tourism resource evaluation is a multi-level and multi-disciplinary comprehensive study, and appropriate scientific and systematic methodology should be adopted to evaluate tourism resources. It can effectively evaluate the quality, type, scale and conditions of the environmental development of the enterprise. The second is the principle of integrity. From the content point of view, the evaluation of tourism resources includes the type, quality, scope of local tourism resources, as well as the socio-economic environment and development conditions, etc., and a multi-level, multi-level systematic evaluation. In the evaluation process, it is necessary to be as comprehensive and complete as possible to meet the integrity requirements. The third is the objective principle. It includes the functionality of the evaluation process and the functionality of the evaluation results. Understand the nature and scope of local tourism resources, select the content and methods of evaluation, and process the data to obtain more scientific and reasonable results. On the other hand, the evaluation results of local tourism resources must be achievable. It must be able to transform the evaluation results into the development and improvement of local tourism resources that combine theory and practice, otherwise the evaluation work will be reduced to theoretical discourse [8, 9].

The development of urban tourism resources is affected and restricted by many factors such as the resources themselves, socio-economic conditions and the tourism market. Therefore, there are many factors that affect the development of urban tourism resources, and the degree of influence is uneven. According to the degree of its influencing factors, the influencing factors are divided into resources, cities and regions.

The city's tourism resources are formed by the interaction of nature and man-made objective aspects. Due to the differences in geographical environment and socio-economic development, the differences in the tourism resources of various cities are very huge. In terms of scale, the smaller the city, the simpler the economic structure, the larger the city, the more complex the economic structure, and the development of its tourism industry is only an integral part of the local economic development. In addition, tourism is a comprehensive industry, and its industrial relevance is particularly strong. The development of urban tourism should stand at the height of the region and focus

on the integration and development with surrounding cities. Therefore, the competition and cooperation relationship in the development of tourism economy between regions are particularly important [10, 11].

2.3 Application of Comprehensive Weighting Algorithm in Tourism Resource Evaluation and Development

Because it is difficult to obtain absolute quantitative index data, the selection of indicators in many tourism resource evaluation studies is mainly based on qualitative descriptions, which shows that most studies use subjective weighting to determine weights. In view of the inevitable shortcomings of the subjective weighting method, this article, based on previous studies, combined with the principal component analysis method and expert advice, carried out the overall subjective and objective weighting, so that the evaluation of tourism resources can be carried out faster, and provide a certain basis for efficient and reasonable development of resources [12]. First, the weighted average calculation model is used to construct the weight evaluation function model, the formula is as follows.

$$j(u) = \sum v_u * y_u \quad (1)$$

Then, calculate the index weight. Normalize different levels of indicators, and get the weight set. Finally, the element weights of each level of the tourism resource index system are obtained.

$$e_u = j_u / \sum_{u=2}^m j_u \quad (2)$$

$$E = (e_1, e_2, \dots, e_m)^R \quad (3)$$

Among them, v_u is the factor score coefficient, and y_u is the score value of the expert on the same indicator. $e(u)$ is the index of different levels, and E is the weight set.

3 Experimental Research on the Construction of Urban Tourism Resource Evaluation and Development Model Based on Artificial Intelligence

3.1 Experimental Environment

In this experiment, the hardware configuration: CPU is Inter i5 2.8 GHz, memory is 12 GB, and hard disk is 1 TB. Software configuration: operating system is Windows XP, web server is Tomcat8.0, background database is MySQL586.6, development and debugging environment is Eclipse5.2. Network configuration: The bandwidth is 1000 Mbps for wired.

3.2 Experimental Process

In this experiment, with the support of artificial intelligence technology and experimental hardware and software equipment, an intelligent model was built to calculate the evaluation level of tourism resources in City G. Then construct the fuzzy evaluation matrix, and use the intelligent model to calculate the evaluation level of some tourism resource indicators. The tourism resource indicators include resource abundance, environmental capacity, location traffic, and ornamental value. Among them, some of the experimental results are shown below.

4 Experimental Analysis on the Construction of Urban Tourism Resource Evaluation and Development Model Based on Artificial Intelligence

In this experiment, a fuzzy evaluation matrix is constructed, and an intelligent model is used to calculate the evaluation level of tourism resources in City G. And the calculation results are shown in Table 1.

Table 1. Calculating results of tourism resource evaluation grade

Project	Not very good	General	Good	Very good
Resource richness	0	2	5	7
Environmental capacity	1	3	6	8
Location traffic	2	3	6	9
Ornamental value	1	5	7	6

It can be seen from Fig. 2 that in this experiment, the intelligent model is used to calculate the evaluation level of tourism resources in City G. Some tourism resource indicators such as resource abundance, environmental capacity, location traffic, and viewing value are all high in evaluation levels, which are 7, 8, 9, and 6 respectively. It can be seen that with the technical support of artificial intelligence, the intelligent model of this experiment can basically evaluate the tourism resources of a city, and to a certain extent can reflect the current situation of local tourism and the direction to be improved, so as to realize the reasonableness of urban tourism resources. Effective development promotes the development of the local tourism industry.

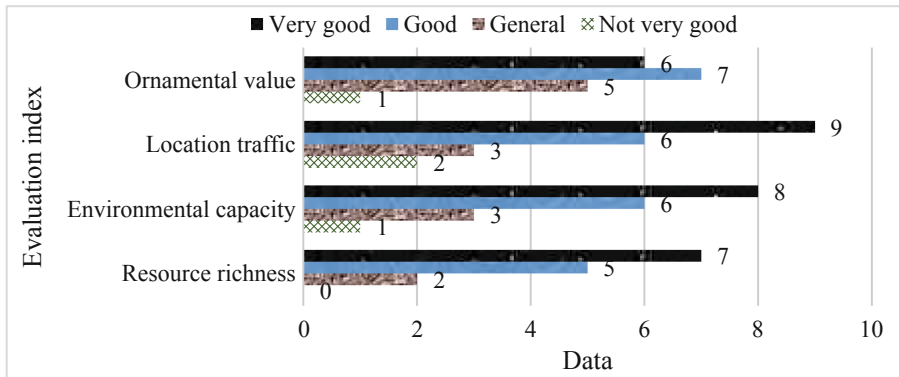


Fig. 2. Calculating results of tourism resource evaluation grade

5 Conclusion

Tourism resources are the basic conditions for the development and development of tourism in a region. The evaluation of urban tourism resources is a qualitative or quantitative analysis of a region's tourism resources in time and space, and predicts its future development trends, providing decision-making support for development and utilization. In order to transform the potential benefits of tourism resources into actual social and economic benefits, it is necessary to analyze the scope, quality, distribution and classification of tourism resources, and conduct scientific and reasonable assessments, and then grasp the focus of local tourism development. At the same time, artificial intelligence has been applied in many fields and has made great contributions in many aspects. Among them, it has also had a huge impact on the development of urban tourism. Therefore, based on artificial intelligence, this article studies the development of urban tourism resource evaluation and development model construction, which is of great significance to the sustainable development of the urban tourism industry.

References

1. Chen, Y.: Research on Economic Efficiency Evaluation of Tourism Cities Based on DEA Model. *Fortune Times*, **171**(08), 187–187 (2019)
2. Yu, H., Qingqing, L., Peng, L.: Regional tourism resource information collection and evaluation of development potential based on ArcGIS technology——taking penglai city as an example. *China Famous City* **000**(012), 47–52 (2017)
3. Chen, J., Xu, S.: The construction of intelligent tourism big data analysis model based on artificial intelligence. *Comput. Knowl. Technol.* **15**(11), 195–196 (2019)
4. Minsung, H., Jason J.J.: Multi-criteria tensor model for tourism recommender systems. *Expert Syst. Appl.* **170**, 114537 (2021)
5. Mishra, R.K., Siddhaling Urolagin, J., Jothi, A.A., Neogi, A.S., Nawaz, N.: Deep learning-based sentiment analysis and topic modeling on tourism during Covid-19 pandemic. *Front. Comput. Sci.* **3**, 775368 (2021). <https://doi.org/10.3389/fcomp.2021.775368>
6. Eleanor, E.C., Cathy Urquhart, M., Claudia T.D., Timothy, J.: Developing augmented reality business models for SMEs in tourism. *Inf. Manag.* **58**(8), 103551 (2021)

7. Zentner, H., Spremic, M.: Typology of digital business models in tourism. *Int. J. E Serv. Mob. Appl.* **13**(2), 21–42 (2021)
8. Cayèré, C., et al.: Multi-level and multiple aspect semantic trajectory model: application to the tourism domain. *ISPRS Int. J. Geo Inf.* **10**(9), 592 (2021)
9. Mohd Norhisham, et al.: Indigenous food recognition model based on various convolutional neural network architectures for gastronomic tourism business analytics. *Inf.* **12**(8), 22 (2021)
10. Leibnitz-Pavel, R.-B., Hugo A.-S., Núñez-del-Prado, M., Joseph, C.G.: Tourism Mobility Modeling based on Credit and Debit Card Transactions. *CLEI 2021* pp. 1–8
11. Boerge, H. J.: Establishing A Basis for Decision Support Modelling of Future Zero Emissions Sea Based Tourism Mobility in the Geiranger Fjord Area. *ECMS 2021*, pp. 88–94
12. Valentinus, R.H., Uwe, S., Victor V.K.: A Tourism Knowledge Model through Topic Modeling from Online Reviews. *ICCDE 2021*, pp. 87–93



Image De-rain Algorithm in the Architecture of the Big Data Platform of the Internet of Things

Yongjun Qi¹, Haiyan Wu^{2(✉)}, and Chadi Altrjman³

¹ Faculty of Megadate and Computing, Guangdong Baiyun University, Guangzhou 510450, Guangdong, China

² School of Management, South China Business College, Guangdong University of Foreign Studies, Guangzhou 510545, Guangdong, China

³ Near East University, Istanbul, Turkey

Abstract. With the continuous development of the Internet of Things, the data security and privacy protection issues of the Internet of Things are becoming more and more noticeable. This article mainly introduces the application of the image rain removal algorithm in the big data platform architecture of the Internet of Things. In order to design multiple small files generated by multiple sensors, we used the IndoorSTG indoor RFID data generation tool to simulate the data generated by 1,000 sensors. In order to increase the calculation speed and save the calculation cost, this paper randomly selects 9100 image pairs regardless of whether it is raining or not, and then randomly retrieves the images. As a network training group, 30,000 image patch pairs will be generated regardless of whether it is raining or not. In the network training process, the Xavier weight initialization method is used, and the Adam optimization algorithm is used. The batch size is 32, the learning rate is set to 0.0005, and a total of 175,000 repeats. DMFS has an average 60% higher write throughput when facing small files than Redis. The results show that the image rain elimination algorithm can improve the stability of the IoT big data platform architecture.

Keywords: Image rain removal algorithm · Internet of Things · Big data · Image decomposition

1 Introduction

In the information society, the development of computer technology is becoming more and more intelligent. The main sources of human access to information are vision and hearing, in which visual information is much larger than auditory information. Human access to information through vision accounts for more than 70%, while only about 10% is obtained through hearing. Computational vision is a kind of computer intelligence that simulates human vision by camera and computer. Computer vision is widely used in many fields, such as remote monitoring, intelligent transportation, target tracking and detection, sports reporting, public security, military and national defense. In the

time frame, the rain lines in the image will be highly aliased based on the background information. This is because separation is difficult, so the background information will be blurred or deleted during the rain removal process, which affects the quality. Rain is similar to the texture of the image and exists in the high-frequency part of the image. Therefore, if the rain in the high frequency part of the image is removed, the rain removal problem is simplified, which helps maintain the background information of the image.

The data requirements in the big data environment are very high, with tens of thousands of read and write requests frequently occurring every 1 s. The performance of relational database is not enough to achieve high simultaneous reading and writing [1, 2]. There are hundreds of millions of data read and written in relational databases. Regarding data, query performance is low, and secondly, the horizontal scalability is low, and it is impossible to collect more than 1000 nodes [3]. In order to recover from failures, RAMCloud saves logs and data backups in ordinary hard drives. Memcached divides the memory into boards of various sizes [4, 5]. These boards are used to store small pieces of string or object data [6]. However, in-memory database cannot be performed. It is now used as an independent data management program [7]. If the network is too deep, the convergence of the network during the training process will become very difficult [8]. Due to the disappearance of the gradient, network training tends to fail [9]. In addition, after the network reaches a certain depth, the performance brought by the deep network will be greatly reduced. After saving the data in the form of a file to a permanent device, in order to achieve high-performance data retrieval services, it is necessary to create an index in the data in the file [10].

On the Internet of articles, in order to perform statistical analysis when obtaining emergency situations, it is necessary to combine the history data for effective information screening. This requires efficient online analysis and processing. In the face of large-scale data ratios, the calculation time of the algorithm for the data aggregation calculation task must be linear or approximately linear or approximately linear, and it is necessary to redesign the statistical analysis algorithm of the conventional database. In the physical Internet environment, optimizing algorithm performance faces considerable challenges. The “Veracity” function based on IoT big data cannot effectively apply traditional statistical algorithms based on data scanning to OLAP queries of IoT big data.

2 IoT Big Data Platform Architecture

2.1 Image Rain Removal Algorithm

The ultimate goal of image processing algorithms is to achieve image rich details and clear contour edges. This is very convenient for future research and use. There are two types of image evaluation: subjective evaluation and objective evaluation. Among them, subjective evaluation is the direct judgment of human beings with the naked eye, which is the subjective classification of picture quality. However, computer vision algorithms are often constrained by the noise of images or videos collected in nature, and can not process the information in the data quickly and accurately. A good data source can make us get the desired information faster and more accurately, which reflects the importance of data preprocessing. The noise of videos or images collected in nature is often caused by bad

weather, and bad weather will increase the difficulty of video or image preprocessing, resulting in many research works on videos or images are limited and can not be applied to all aspects of our life. Therefore, the basic research on video or image preprocessing collected in bad weather is particularly important.

Assuming that the input of the PolyInception module is x , the nonlinear mapping of the Inception residual part in the module can be expressed as:

$$H(x) = W_3 \text{Concat} \left(\sigma(W_1 x + b_1), \sigma \left(W_2^3 \sigma \left(W_2^1 x + b_2^1 \right) \right) \right) + b_3 \quad (1)$$

Among them, W_1 and b_1 respectively represent the weight and bias of the 1×1 Conv(64) branch.

The output of the nonlinear mapping of the 2-way module can be expressed as:

$$F_{2\text{-way}} = \sigma(H(x) + H(x) + x) \quad (2)$$

The output of the nonlinear mapping of the poly-3 module can be expressed as:

$$F_{\text{poly-3}} = \sigma(H(x) + H(H(x)) + H(H(H(x)))) + x \quad (3)$$

File query performance optimization strategy. The optimization of file query performance mainly depends on the disk I/O times required for query. Placing files with stronger access relevance into a large file can effectively reduce the time overhead of query. For different query types, different access Association evaluation models and different clustering algorithms need to be used. This paper designs an access relevance evaluation model and clustering algorithm for object point query, the most common query type in the Internet of things application scenario. Archive file is a file archive policy that combines files and puts them into HDFS blocks. It works by creating a hierarchical file system on HDFS. It is created by the archive tool from a collection of files. The sequence file uses the file name as the key and the file content as the value. The file is written into the sequence file through parallel program, and the small file is indexed in the form of binary tree. Although the method of creating archive file and sequential processing small file does reduce the memory utilization of name node.

2.2 Big Data of the Internet of Things

Data is sent to the data center in the form of a stream, and the data arrives very fast, especially on the Internet. In order to meet the timeliness of the query, the index construction speed consistent with the data retrieval system is required. The results of data retrieval are very different from the results of the next second, and there is a high requirement for the index construction speed of the data retrieval system. It is generally believed that if a large amount of data is used to make an index, especially in the case of the main memory configuration index, due to the capacity limitation and high cost of the main memory, the space overhead is very high. The output characteristics of the first layer and all bottleneck structures in the network will be passed to the subsequent layers. There is a direct positive connection between any two layers in the network, forming a tightly connected network structure. This structure can strengthen the dissemination of network

information. Each layer in the network can use the features extracted from the previous layer. This helps the high-frequency prediction subnet learn more detailed information about no rain.

In terms of big data storage of IOT perception, a large number of sensor devices collect and upload data in real time. The storage system needs to store a large number of perception in near real time, that is, it has high requirements for write throughput. Due to the temporal and spatial characteristics of the perceived data of the Internet of things, the perceived information includes time information and location information. The location information changes with time. The existing storage structure only records each data directly and does not reflect its temporal and spatial characteristics, which brings difficulties to rapid query and analysis. In the aspect of big data query perceived by the Internet of things, due to the spatio-temporal and multidimensional characteristics of the perceived data of the Internet of things, the traditional key value data storage system is difficult to carry out efficient multi-dimensional query and analysis. In multidimensional query, a large amount of data needs to be scanned, which takes a long time, and even affects the normal query and write operations. IOT sensors will be popularized in all fields and our lives, forming an ubiquitous IOT world and providing users with convenient and efficient services. How to deal with the high-throughput writing and efficient query of big data perceived by the Internet of things has become a core problem to be solved.

3 Simulation Experiment of IoT Big Data Platform Architecture

3.1 Experimental Environment

In order to design multiple small files generated by multiple sensors, we used the IndoorSTG indoor RFID data generation tool to simulate the data generated by 1,000 sensors. The principle of IndoorSTG is to simulate the movement of physical objects based on semantics. RFID sensor data can be simulated and generated by configuring indoor structure, sensor location and other parameters. Each RFID sensor continuously detects moving objects, and when the amount of data reaches a certain threshold, the file data is sent to the storage system for storage. There are two main criteria for evaluating the data set: the first is the degree of conformity with the real data. The better the conformity with the real data, the closer the features learned by the machine are to the real distribution, and the better the effect when the model is applied to the actual task; The second important factor is the size of the data set itself. In the training process of the deep learning model, if the training set is too small, the model will inevitably fall into over fitting, that is, the model fits well on the training set, but the effect on the test set is very poor. This happens mainly because the data set itself is too small to represent a wide range of situations statistically.

3.2 Network Training and Testing

Firstly, the network camera directly collects sequence images. This part of data is mainly represented by some static background images in the data set, such as buildings, streets without moving objects, bright flowers and street statues. This is mainly because the rain

and rain free images in these scenes will not be affected by moving objects, So we can try to make the training data conform to the actual situation. The second way is to directly crawl the rain free pictures through the network. The data obtained in this way is mainly shown in some pictures with dynamic objects in the data set, such as cars on the street, football teams on the grass, landscape pictures and night scenes of some streets. Because we will simulate and add the effect of rain to these pictures, Therefore, there is no need to pay attention to whether there are moving objects in their scene. The data we collected before can not fully meet the training requirements, which mainly has the following defects: first, the amount of data is still insufficient. Less than 20000 pictures are used as the training set, which is easy to make the training fall into over fitting, so the data needs to be expanded; Second, many of the collected pictures do not meet the requirements, mainly reflected in the low resolution, affecting the recognition, the watermark blocking the picture itself, and the contrast is too strong to simulate rain; Third, the size of the pictures crawled on the Internet is different, which is not conducive to the network to find the statistical distribution of features during training, and the final results of pictures of different sizes in the designed network will be different, which is also not conducive to the design of the model, so it is necessary to normalize the obtained pictures. In order to increase the calculation speed and save the calculation cost, this paper randomly selects 9100 image pairs regardless of whether it is raining or not, and then randomly retrieves the images. As a network training group, 30,000 image patch pairs will be generated regardless of whether it is raining or not. In the network training process, the Xavier weight initialization method is used, and the Adam optimization algorithm is used. The batch size is 32, the learning rate is set to 0.0005, and a total of 175,000 repeats.

4 Experimental Results

Create 10, 20, 30, and 40 tuples for each object to simulate the uncertainty of the data. In the same object, the tuples are all located in a multi-dimensional orthogonal area, and the center of the area is the original data. Multidimensional data. For each dimension, the length of the region is set to 0.5%, 1%, and 2% of the value range. In such a field, the instance is distributed according to a multi-dimensional Gaussian, and its expected value is the center of the area. The SSIM comparison of the six single-image rain removal algorithms on the synthetic rain map test set is shown in Fig. 1. The PSNR comparison of the six single-image rain removal algorithms on the synthetic rain map test set is shown in Fig. 2. The time overhead of convolution-based aggregation increases with the increase of spacebudget. This is due to the longer pmf length in the higher-level Cuboid aggregation operation, which increases the time overhead of convolution, and the aggregation based on Sketch the cost of partial materialization of the strategy increases linearly. Compared with convolution, the time cost is greatly reduced. For task performance, the error rate is less than 13%, and the power performance error is within an acceptable range. The goal of energy efficiency optimization in this paper is to perform all submitted stream computing tasks and minimize the energy consumption of the cluster without reducing the performance. Among them, two task forms on yarn are considered, namely batch task and online task. Batch tasks are submitted to yarn at the same time. For example, yarn is used to analyze sales data during business hours, or to filter and analyze data

sets collected from websites. This type of tasks are submitted to the computing cluster periodically and run in batches. Batch tasks are widely used in many traditional business environments that need offline processing services. Corresponding to batch tasks, online work is submitted to the computing cluster irregularly one by one. This task submission method is common in public cloud platforms.

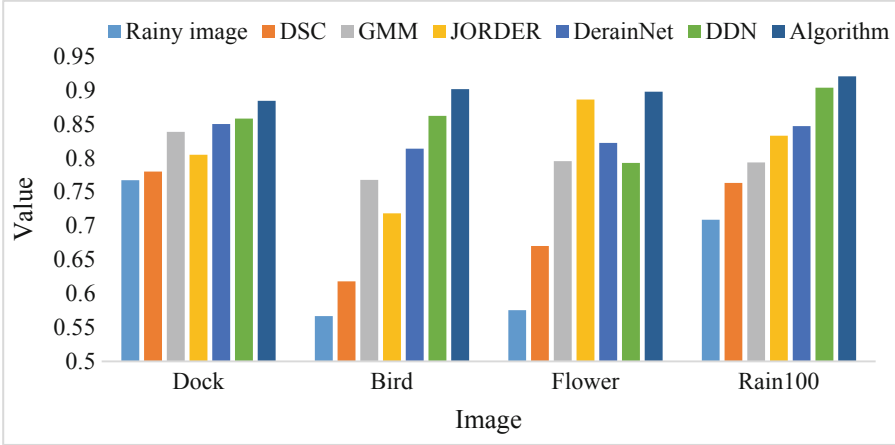


Fig. 1. SSIM comparison of six single-image rain removal algorithms on the synthetic rain map test set

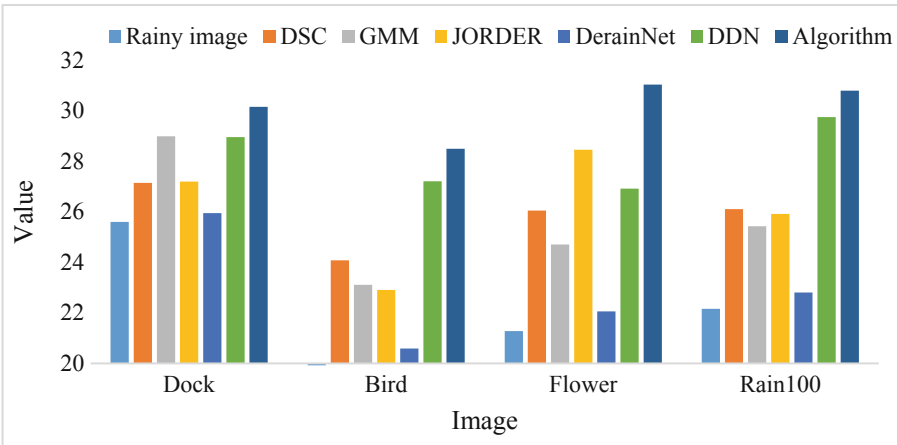


Fig. 2. The PSNR comparison of six single-image rain removal algorithms on the synthetic rain map test set

Part of the materialization tests the trade-off between space cost and material performance, and inquires about changes in performance under various storage costs. Spatial tools are used to represent the storage cost of part of the material. Here, the space ejector is the ratio of the space overhead of the partial materialization and the total cost

materialization. The memory write rate comparison is shown in Table 1. Experimental results show that the write throughput of DMFS for small files is 60% higher than that of redis on average. This is mainly because DMFS uses multi-threaded concurrent write technology to improve the write throughput of the system. The threading architecture of Redis is single-threaded. DMFS is better than Redis. The performance of memcached has increased by 407%, mainly due to the serious locking conflicts caused by write operations in memcached. The performance of DMFS has increased by 625% and 182% respectively over spark and tachyon. Firstly, the image is divided into high-frequency part and low-frequency part by guided filtering. Secondly, by appropriately enhancing the low frequency as the reference image and using the high frequency part as the input image, the rain and snow are removed from the high frequency part, and then the high frequency part without rain and snow is added to the low frequency part to obtain the initial restored image. Then, the valley edge in the image can be restored by taking the minimum value of the initial restored image and the input image to obtain a clear image. Finally, the balance between the clear image and the initial image is taken as the guide image, and the guide filter is used again for the clear image to obtain the final restoration result.

Table 1. Memory write rate comparison

	DMFS	Redis	Memcached	Spark	Tachyon	RAMCloud
10	1604.67	614.85	2690.14	1445.32	2385.05	2691.93
20	1987.96	2121.60	586.77	2751.87	2634.09	698.62
30	1541.24	1276.80	1284.92	734.24	1894.13	989.91
40	2162.55	2014.87	1559.31	2589.27	2919.50	875.13
50	1053.11	2078.01	966.03	1962.68	766.28	1014.33
60	2503.59	2159.79	2448.34	1171.50	1081.58	1803.48
70	2543.17	1757.14	1717.82	1779.08	1468.74	2286.52
80	1483.63	1552.31	1849.80	773.79	2803.75	2486.09
90	2177.97	1151.84	2490.88	576.14	2379.37	2341.34
100	902.53	1481.41	1189.08	2164.17	2160.93	910.39

5 Conclusions

From the perspective of time domain, this paper analyzes the variation characteristics of raindrop brightness value, uses the improved k-means clustering method to detect raindrops, and then uses the color characteristics of raindrops to realize the initial optimization. Finally, the removal of raindrops in video images is realized by statistical characteristics. Compared with the traditional rain removal algorithm, the algorithm in this paper has better rain removal effect, and the restored image details are clearer. For

the energy consumption optimization of streaming task scheduling framework, only the power consumption of CPU is optimized at present, but the power consumption characteristics of memory, bandwidth, disk and other hardware need to be analyzed in the future. Through the analysis of relevant theories, this paper believes that the construction of local manifold structure in the extraction process of refined rain and snow map can greatly reduce the loss of inherent information in the image. Rain and snow destroy the inherent information and internal laws contained in the image, so accurate extraction of rain and snow stripe information can avoid the secondary destruction of image content by image processing algorithm. Starting from the manifold learning theory, this paper uses the low-dimensional embedded manifold to describe the inherent information of the image, then uses the global information and local information of the image to solve the refined rain and snow map of the image, and finally applies it to the rain and snow removal of the image.

Acknowledgements. Feature innovation project of colleges and universities in Guangdong province: 2020KTSCX163.

Feature innovation project of colleges and universities in Guangdong province: 2018KTSCX256.

Guangdong baiyun university key project: 2019BYKYZ02.

Special project in key fields of colleges and universities in Guangdong Province: 2020ZDZX3009.

References

1. Wu, Q., Wang, L., Ngan, K.N., Li, H., Meng, F., Xu, L.: Subjective and objective de-raining quality assessment towards authentic rain image (2019)
2. Arjun, K.P., Sreenarayanan, N.M., Kumar, K.S., Viswanathan, R.: Emerging IoT-big data platform oriented technologies (2020)
3. Yang, H.F., Chen, C., Chen, K.: Using big data analytics and visualization to create IoT-enabled science park smart governance platform. In: HCI in Business, Government and Organizations. Information Systems and Analytics (2019)
4. Zubair, N., Niranjana, A., Hebbar, K., Simmhan, Y.: Characterizing IoT data and its quality for use (2019)
5. Rajiv, R., Omer, R., Surya, N., et al.: The next grand challenges: integrating the Internet of Things and data science. *IEEE Cloud Comput.* **5**(3), 12–26 (2018)
6. Atlam, H.F., Alenezi, A., Hussein, R.K., et al.: Validation of an adaptive risk-based access control model for the Internet of Things. *Int. J. Comput. Netw. Inf. Secur.* **1**(1), 26–35 (2018)
7. Du, L., Du, Y., Li, Y., et al.: A reconfigurable streaming deep convolutional neural network accelerator for Internet of Things. *IEEE Trans. Circuits Syst. I Regul. Pap.* **65**(1), 198–208 (2018)
8. Papert, M., Pflaum, A.: Development of an ecosystem model for the realization of Internet of Things (IoT) services in supply chain management. *Electron. Mark.* **27**(2), 175–189 (2017). <https://doi.org/10.1007/s12525-017-0251-8>
9. Lyu, X., Tian, H., Jiang, L., et al.: Selective offloading in mobile edge computing for the green Internet of Things. *IEEE Network* **32**(1), 54–60 (2018)
10. Reka, S.S., Dragicic, T.: Future effectual role of energy delivery: a comprehensive review of Internet of Things and smart grid. *Renew. Sustain. Energy Rev.* **91**, 90–108 (2018)



Intelligent Analysis Technology of Sports Training Posture Based on Deep Learning

Cong Zhou^(✉)

Department of Social Training, Shandong Polytechnic College, Jining, Shandong, China

Abstract. Sports training is a complex, challenging and uncertain process. It needs to consider many factors, such as speed, acceleration, etc. For sports training, posture control is one of its core issues. In order to improve the athlete's sports skills and posture, it is necessary to conduct research on the intelligence analysis of sports training posture. This article mainly applies analytical and experimental methods to intelligently analyze and detect sports training postures, and proposes corresponding intelligent analysis methods. The experimental results show that under the same FLOPs, the global network has better results than the one-stage hourglass network, which can reach 66.8% .

Keywords: Deep learning · Motor training · Pose recognition · Intelligent analysis

1 Introduction

Sports training is a kind of athletic technology that is based on the body and achieves a certain target effect through the perception and analysis of external stimuli. In the field of sports, sports training has become a behavioral activity that strengthens athletes' skills and strengthens their limbs. The main goal of sports training is to improve its posture and physical abilities. In practical applications, the motion trajectory tracking technology can accurately control the height and angle.

There are many theories for the research on the intelligent analysis technology of sports training posture based on deep learning. For example, some researchers have proposed that human gesture recognition is widely used as a research hotspot in computer vision and pattern recognition in many fields such as human-computer interaction, intelligent video surveillance, analysis, sports, and medical assistance [1, 2]. In addition, it is also said that motion capture technology, as an important branch of natural multi-modal human-computer interaction technology, has developed rapidly in recent years [3, 4]. In addition, in order to improve the accuracy of the athlete's posture analysis, some people quantify the athlete's video data for training feedback, use image processing technology to process the athlete's video, and provide a multi-functional optical flow analysis model

that tracks human commonality [5, 6]. Therefore, in order to solve the current problem of inaccurate sports posture testing, this article studies the intelligent analysis technology of sports training posture.

This article first studies the human posture analysis of deep learning. Secondly, it briefly describes the motion training gesture recognition in deep learning. Then it analyzes the moving target detection method. Finally, an experimental test was carried out, and the data results were obtained.

2 Intelligent Analysis Technology of Sports Training Posture Based on Deep Learning

2.1 Human Posture Analysis Based on Deep Learning

(1) Deep learning

Deep learning is a method of processing training data on a computer based on artificial brain networks, while sports training needs to be achieved by obtaining relevant information in real time.

Its appearance has led to the rapid development of speech recognition, image processing, and natural language processing. Therefore, deep learning is becoming more and more popular in academic theories and industrial applications. Combining deep learning technology to solve the problem of classification and detection of motion status. To this end, it is necessary to understand the related technologies of deep learning. Deep learning is a wide range of machine learning technology and architecture, which is characterized by the use of hierarchical non-linear information processing methods, which are hierarchical in nature. Among the monitored deep learning networks, there are mainly LSTM network models and neural network models such as CNN. The deep learning network of the hybrid system mainly includes a combination model of deep belief network and deep neural network, a new model combining deep belief network and Markov model, etc. [7, 8].

Neural network and mechanical structure analysis method. Neural network has a variety of classification capabilities and different types of neuron unit composition. Its biggest feature is that its nonlinearity, time-varying and unsupervised characteristics make it easy to understand and can be used in deep learning; in addition, because artificial cells are working. In the process, new stimuli are constantly produced, so there are certain requirements for in-depth training [9, 10].

(2) Artificial intelligence

The field of artificial intelligence has developed vigorously in the past few decades and has developed into a future-oriented theme with increasingly perfect theoretical foundations, universal application possibilities and a high degree of interactivity. With the rapid development of data generation and hardware and software technology, artificial intelligence has become widespread. In short, artificial intelligence is to let computers think and automatically process tasks like humans; because the concept of artificial intelligence is very extensive, more technical and theoretical research is needed [11, 12].

(3) Machine learning

Machine learning is a discipline dedicated to the study of how to use computational methods and use the acquired knowledge to improve the performance of the system itself. In computer systems, the so-called “experience” usually exists in the form of “data”, so the content of machine learning research is to use the characteristics of the data to generate the algorithm of the model. From the core perspective of machine learning, optimization and statistics are the two core supporting technologies. Machine learning can be seen not only as a method, but also for pattern recognition or data mining. There are four main types of problems in the field of machine learning: prediction, clustering, classification, and dimensionality reduction. At present, in many fields of computing science, including graphics and image processing, software engineering, computer vision, and natural language processing, machine learning can serve them.

(4) Pattern recognition

With the advent of the big data era, deep learning has been successfully applied in the field of computer vision. Therefore, studying how to use deep learning to solve the problem of human pose estimation is another focus that researchers in the field of human pose estimation based on graphic structure models should study. The first step is to obtain an approximate human pose in the input image, and then feed the result back to the same deep convolution grid, and repeat the loop until the exit condition of the grid is met. This model converts the human body posture problem into a regression problem, and obtains a regressor that estimates the human body posture from the overall situation of the human body through training and learning. Compared with the classic paired model method, this method significantly improves the accuracy of human pose estimation.

Human pose estimation methods based on model comparison have some applications in static images or videos. Due to the diversity of human postures, it is difficult to list all possible human posture models in practical applications, so this type of model is only suitable for the analysis of a few specific actions. This model is currently the most widely used model, and it is also a research topic in the field of human pose estimation. The graphic structure model decomposes the human body into some interconnected common points. It consists of three parts: the appearance model of the parts, the geometric constraint relationship between the parts and the graphic alignment. The model not only uses Markov network for modeling, but also uses fast and efficient graphical thinking, which can flexibly represent any human posture.

2.2 Sports Training Posture Recognition in Deep Learning

Deep learning means that in a movement process, a certain goal can be achieved by operating multiple objects, but this technology is based on humans’ accumulated experience and knowledge. And what we generally think of as sports training mainly refers to a kind of highly difficult, highly complex action in a large range, which is a non-linear programming design. For us, deep neural network applications and sports science can

be expressed as: First, it must be able to determine an optimal output point. Secondly, it is necessary to calculate each neuron in the network, and then find the optimal solution or output result based on the input data.

In sports training, the concept of deep learning is introduced in the traditional sense, and a new height is obtained by reprocessing the initial position information and current posture data. However, the original image is often composed of a large number of points and has the same characteristics. Therefore, when it is necessary to obtain certain information, errors or omissions may occur, which may cause the output result to be different from the actual situation, even excessive deviation or too large inaccuracy, causing unnecessary losses. Deep learning provides a new method to solve kinds of series of the questions.

2.3 Moving Target Detection

The purpose of moving target detection is to achieve target extraction and segmentation. That is, the image sequence is processed, and the target of interest is extracted based on the geometric or statistical characteristics of the target. Effectively extracting human targets or features is an indispensable step for pose estimation. Therefore, the research focus of this chapter is to detect and extract human features from video or sequence images for subsequent human pose estimation and the construction of auxiliary training systems.

The success of an analysis algorithm depends to a large extent on the characteristics it uses and defines. Color features, texture features, shape features, edge gradient features, spatio-temporal relationship features, etc. are image features that are frequently used. The color function is one of the most commonly used functions in image functions. The color attribute describes the pixel attribute of the image area or the surface attribute of the scene corresponding to the image. As an overall attribute, texture attributes and color attributes have the same mechanism for characterizing attributes. The difference between the two is that the texture attributes are based on several pixel attributes. Both color features and texture features have their limitations. Due to the influence of external factors such as light, shadow, clothing, etc., the feature extraction in the image will be greatly affected.

(1) Background subtraction

The background subtraction method can be regarded as a special screen difference method. Once the background model is established, the approximate shape of the foreground object can be obtained by comparing the current image with the background model and subtracting the known background information. The background image can be represented by $Y(a, b)$, the image sequence is represented by $X(a, b, m)$, the position coordinates of the image are represented by (a, b) , the number of images is represented by m , and the difference image can be represented by any A kind of expression to remove the noise in the foreground image caused by the background or illumination changes. The specific formula is:

$$Q(a, b) = \begin{cases} 1, & (X(a, b) - Y(a, b)) \geq R \\ 0, & (X(a, b) - Y(a, b)) < R \end{cases} \quad (1)$$

The method of determining the background pixel is: the pixel gray value of the current frame image is subtracted from the pixel gray value of the background image, and the difference between the two is within a certain threshold range; similarly, we can determine that the pixel has a human body. What happens is that the difference between the pixel points of the current frame image and the pixel points of the background image is not within the threshold range.

(2) Inter-frame difference method

The inter-frame difference method is a kind of background subtraction method. Its background model is the image of the previous frame. The detection idea of the frame difference method is to detect the changed areas in two adjacent frames of images, and then obtain the approximate outline of the target. When the environment brightness is basically unchanged, if a moving object appears in the camera's field of view, the interval between two adjacent frames is short, and the motion information in the image is extracted by binarizing the gray-scale difference image of the adjacent frame. The specific formula is:

$$Q_I(a, b) = |X_I(a, b) - X_{I-1}(a, b)| \quad (2)$$

The change between the grayscale image of frame I $X_I(a, b)$ and frame $I - 1$ $X_{I-1}(a, b)$ is represented by a difference image.

3 Human Body Gesture Recognition Experiment

3.1 Data and Training

This article mainly uses MSCOCO's detection data, which has a total of 50 categories and more than 1000 pictures. In the training phase, ResNet101 is used as the backbone network, and the ROIAlign method takes the maximum value. At the same time, FPN is used for the fusion between different feature layers. When generating labels, the overlap rate with the label frame is greater than 0.6 as a positive example. The ones less than 0.2 are regarded as negative examples, and the others are not regarded as samples. Soft-NMS is used for non-maximum suppression, and the threshold is set to 0.6. In order to improve the performance of the network, there are five anchor scales: 3, 6, 9, 12, and 15. The aspect ratio is 1, 2, and 4. The number of online difficult sample mining is set to 32, and the weight attenuation coefficient of $1e - 4$ and the momentum coefficient of 0.8 are used. In terms of optimization, this article uses a segmented learning strategy. It is worth noting that the initial learning rate used here is related to the number of batches per GPU. The larger the number of batches, the greater the initial learning rate. In the testing phase, this article uses a common non-maximum suppression threshold of 0.5.

3.2 Data and Evaluation Indicators

MSCOCO human body key point data set MSCOCO's training set has 30K images, at the same time there are 100K human bodies, and the verification set has 5K images.

The test set contains test-dev and test-challenge, each containing 15K pictures, and its evaluation standard is mAP based on OKS, where OKS is defined as the similarity between different people’s postures.

Posetrack is mainly a data set proposed for posture tracking in videos. It has 200 videos, including more than 15K video frames and more than 100K personal posture annotations. Among them, there are 25 sections in the training set, 25 sections in the verification set, and the test set. There are 150 segments. It can be seen that each picture in the Posetrack data set contains more human body annotations, which poses a certain challenge to the algorithm of this article. The evaluation standard is based on the mAP of PCKh. The PCKh here is similar to the above OKS. It is used to measure the distance between two human postures. At the same time, there are three different subtasks based on the Posetrack data set, namely: 1) human pose estimation of a single frame picture; 2) human pose estimation based on multiple frames; 3) multi-person pose tracking.

3.3 Experimental Process

Data augmentation is extremely important for realizing the scale invariance and rotation invariance of the algorithm. Therefore, here, this article first trims the detected human frame while maintaining its own aspect ratio. In the ResNet50 network, the crop size of this paper is 186×192 , and for the ResNet101 network, this paper uses a crop size of 286×288 , and only the 100 candidate boxes with the highest scores obtained by the detection are operated to reduce the amount of calculation. On this basis, this paper uses random scales with random rotations and horizontal flips for data amplification. In order to minimize the variance of the prediction results, this paper uses a 2D Gaussian kernel on the predicted heat map. This article not only predicts the key point position of the test picture, but also predicts the key point of the image that is flipped left and right, and takes the average of the two as the final prediction result.

4 Analysis of Experimental Results

4.1 Method Comparison

This paper compares the proposed CPN network with the cascaded hourglass algorithm with different stages, the ResNet50 network with different dilated convolutions, and the CPN network without online hard sample mining technology. The results are shown in Table 1:

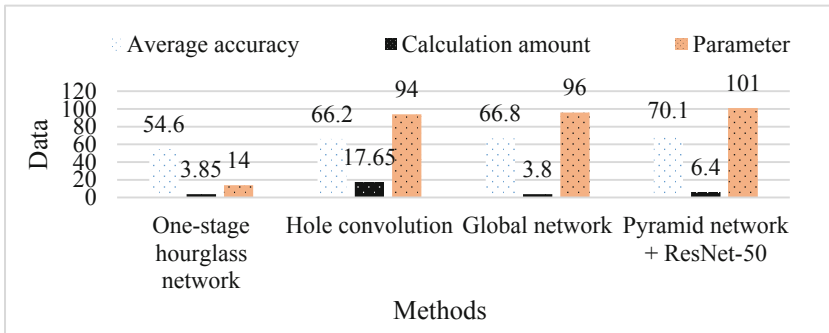
Table 1. Comparison of other methods of cascaded pyramid network

	Average accuracy	Calculation amount	Parameter
One-stage hour glass network	54.6	3.85	14

(continued)

Table 1. (continued)

	Average accuracy	Calculation amount	Parameter
Holecon volution	66.2	17.65	94
Global network	66.8	3.80	96
Pyramid network+ResNet-50	70.1	6.4	101

**Fig. 1.** Comparison of other methods of cascaded pyramid network

As shown in Fig. 1, we can see that CPN has a better ability to compromise speed and accuracy. The global network has better results than the one-stage hourglass network.

4.2 Results of Attitude Tracking

This paper uses the MOT16 standard data set to verify the effectiveness of the proposed algorithm, and then tests the effect of human posture tracking in the video on the Posetrack data set. The details are shown in Table 2:

Table 2. Results of posture tracking

	MOTA	MOTP	FAF
SOT+MOT	43.5	74.6	2.3
MHT	42.1	75.5	1.1
NOMT	46.2	75.6	1.8
JMC	46.1	74.7	1.2

As shown in Fig. 2, we can know that because the data set has very strict labeling standards, the labeling results are very reliable and can better reflect the quality of the

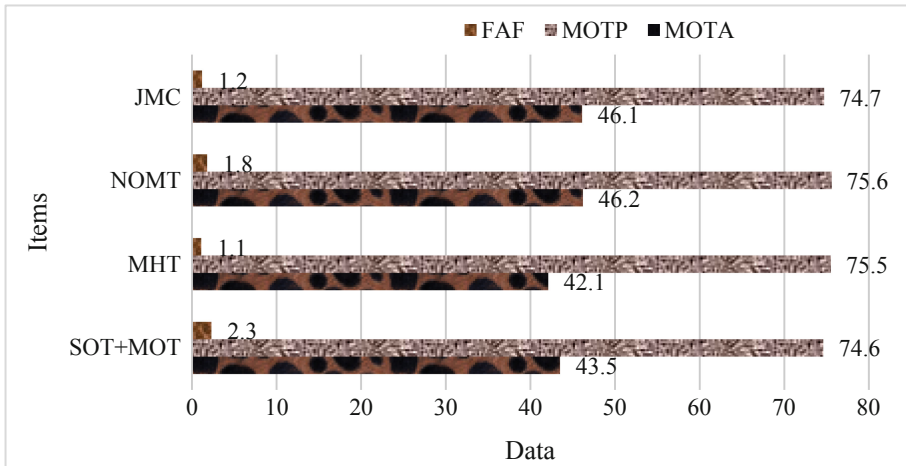


Fig. 2. Results of posture tracking

detector. The comparison between the method in this paper and other methods on MOT16 shows its superiority.

5 Conclusion

In order to overcome the shortcomings of traditional structured probabilistic neural networks in learning, the deep brain network is used to provide a training space, and the deep neural network is designed to realize the analysis and research of the motion model. Sports training is in a specific environment, through some simple methods, to make certain objects reach a certain goal, or a certain technique or rule. Sports training is mainly used to solve the need to do some corresponding actions when humans have a certain degree of difficulty. Through the understanding of deep learning and sports training posture analysis, the relevant algorithms and systems proposed in this paper can be effectively tested and the training effect can be obtained.

References

1. Sergis, S., Sampson, D.G., Pelliccione, L.: Investigating the impact of Flipped Classroom on students' learning experiences: a self-determination theory approach. *Comput. Hum. Behav.* **78**, 368–378 (2018)
2. McNally, B., et al.: Flipped classroom experiences: student preferences and flip strategy in a higher education context. *High. Educ.* **73**(2), 281–298 (2016). <https://doi.org/10.1007/s10734-016-0014-z>
3. Jovanovic, J., Gasevic, D., Dawson, S., et al.: Learning analytics to unveil learning strategies in a flipped classroom. *Internet High. Educ.* **33**, 74–85 (2019)
4. Benzaid, C., Lounis, K., AlNemrat, A., et al.: Fast authentication in wireless sensor networks. *Future Gener. Comput. Syst.* **55**, 362–375 (2016)

5. Kai, Z.: Design and data analysis of a wearable swimming posture measurement system based on the Internet of Things. *J. Test. Technol.* **033**(002), 131–137 (2019)
6. Sohn, K., Kim, Y.Y., Saxena, N.: A new cross-layer QoS-provisioning architecture in wireless multimedia sensor networks. *KSII Trans. Internet Inf. Syst.* **10**(12), 5286–5306 (2018)
7. Borawake-Satao, R., Prasad, R.: Comprehensive survey on effect of mobility over routing issues in wireless multimedia sensor networks. *Int. J. Pervasive Comput. Commun.* **12**(4), 447–465 (2017)
8. Millar, A.: Perceptual recognition and strange environments: reply to Broncano-Berrocal. *Proc. Aristotelian Soc. (Hardback)* **117**(2), 207–214 (2017)
9. Famila, S., Jawahar, A., Sariga, A., Shankar, K.: Improved artificial bee colony optimization based clustering algorithm for SMART sensor environments. *Peer-to-Peer Netw. Appl.* **13**(4), 1071–1079 (2019). <https://doi.org/10.1007/s12083-019-00805-4>
10. Graziano, A.D., Marchetta, V., Cafiso, S.: Structural health monitoring of asphalt pavements using smart sensor networks: a comprehensive review. *J. Traffic Transp. Eng. (English Ed.)* **7**(5), 639–651 (2020)
11. Banach, M., Wasilewska, A., Dlugosz, R., et al.: Novel techniques for a wireless motion capture system for the monitoring and rehabilitation of disabled persons for application in smart buildings. *Technol. Health Care Official J. Eur. Soc. Eng. Med.* **26**(2), 1–7 (2018)
12. Dash, K.S., Puhan, N.B., Panda, G.: Unconstrained handwritten digit recognition using perceptual shape primitives. *Pattern Anal. Appl.* **21**(2), 413–436 (2016). <https://doi.org/10.1007/s10044-016-0586-3>



Design and Implementation of English Speech Scoring Data System Based on Neural Network Algorithm

Xia Sun(✉)

Department of Public Education, Shandong Polytechnic College, Jining, Shandong, China

Abstract. With the improvement of the performance of the Internet and other mobile terminals, English speech scoring technology based on neural network algorithms has become more and more favored by business development and daily life. Therefore, how to provide a voice scoring data system with high recognition performance and strong robustness has become more and more obvious, and it is also particularly important. The purpose of this paper is to study the design of English speech scoring data system based on neural network algorithm. This article makes full use of the freedom of neural network learning and the adaptability of multiple data domains. This paper has developed a set of data system for scoring spoken English. Students can use the system to simulate film dubbing, and the system can evaluate the multi-index data in the recordings submitted by the students. Experimental research shows that the improved BP algorithm used in this paper has a higher oral English recognition rate than the traditional BP algorithm by about 10%, and the optimized algorithm training performance and recognition rate are significantly better than the traditional algorithm.

Keywords: Neural network · Speech scoring · Evaluation system · Machine learning

1 Introduction

English speech scoring technology based on neural network algorithms can automatically evaluate the testers' spoken language, and the research results of pronunciation quality assessment technology can help students learn and understand English better, and help improve their pronunciation [1, 2]. Especially for second language learners who lack the language environment and professional guidance [3]. In addition, it can also be used for automated evaluation of spoken computer tests, because it is not subject to interference from subjective factors, and can significantly improve the effectiveness and quality of the score.

In the research on the design of English speech scoring data system based on neural network algorithm, many scholars have studied it and achieved good results. For example, Kaur G designed a communication network based on the PESQ algorithm based on the characteristics of the voice quality of the communication network. The scheme of speech quality assessment system [4]. MD Pawar proposed an overall design scheme of a

communication network voice quality assessment system based on the PESQ algorithm, as well as the selection of the main software architecture and hardware [5]. It can be seen that the combination of algorithms and systems to evaluate spoken English ensures that the evaluation results are fast, true and reliable, and can meet people's needs in all aspects. It has broad application prospects in the assessment of spoken English speech quality.

This paper proposes an optimized parameter extraction method based on MFCC. Then summarized the advantages and disadvantages of the traditional BP neural network algorithm in the speech recognition system, and proposed the method of simultaneously increasing the momentum factor and using the adaptive learning rate to optimize.

2 Design and Research of English Speech Scoring Data System Based on Neural Network Algorithm

2.1 Voice Recognition Prototype System

(1) System function

The work content completed by each module is as follows.

- 1) Voice preprocessing. Remove the non-sound segment, and at the same time process the noise segment speech collected in the test set environment into groups, Fourier transform, and intra-group spectrum averaging, and add the noise spectrum to the training set. After the above processing, the mixed training set is obtained as The input of the acoustic model.
- 2) Build an acoustic model. The acoustic model is built, the BP backpropagation algorithm training parameters are used in the early stage, and the improved NWBP algorithm training weights are called in the later stage of the model training. The model completes the mapping of speech features and phonemes or words [6, 7].
- 3) Statement display. Display the relevant information of the sentence corresponding to the input voice.

(2) Key category design

- 1) SoundAdding class: The noise of the test set is obtained through the endpoint detection of the EndpointDetection class, and the spectrum obtained by the iff transformation is added to the spectrum of the training set, and then the time domain map of the noisy speech is obtained by the ifft transformation.
- 2) SpectrogramGet class: Draw a speech spectrogram. The training set with noise added by the SoundAdding class and the test set after the endpoint detection of the EndpointDetection class inherit the PretreatDone class for basic speech processing, and then draw the spectrogram. The SpecColorMap function can draw the spectrogram color board.
- 3) SpeechTraining class: The spectrogram obtained through the SpectrogramGet class is used as the input of the acoustic model, and the backpropagation algorithm is used to train the model. This class is implemented using some methods in the DeepLearnToolbox-master toolkit.

2.2 Error Detection System Based on KT Divergence Distance Metric in Phonetics Space

- (1) Description of the problem of detection of spoken pronunciation errors
 In the language learning database, each phoneme p has a large number of samples, which are divided into two categories according to manual annotation: one is the correct pronunciation sample, and the other is the wrong pronunciation sample [1, 8]. Through the forward propagation of the DNN, they are respectively mapped to two point sets in the phonetic space. If the center point of the point set is used to represent its position in space, it represents the mean value of the posterior probability distribution of the correct pronunciation and the wrong pronunciation of the phoneme p .
- (2) Classifier design
 The task of detecting pronunciation errors for each phoneme is abstracted as a binary classification problem. Assuming that there are M different phonemes in the phoneme set, the traditional classification method divides the entire data set into M parts according to the corresponding phonemes, and learns a binary classifier for each phoneme separately. In this paper, a multi-task classifier based on neural network is proposed, which is different from traditional classification algorithms: M two classifiers are embedded in the same neural network framework, and each classifier shares the underlying neural unit to classify M two classes. Trainers are trained together instead of separately.

2.3 Spoken Language Evaluation Data System Design

- (1) System design
 In the data module, the system visually displays the individual evaluation results of students, the evaluation results of different indicators, the evaluation results of different corpora, and the evaluation results of different classes through diversified forms such as radar chart, bar chart, pie chart, and line chart., From different angles, assist teachers, find common problems in daily teaching, and provide data support for personalized teaching.
 Spoken language learning anytime and anywhere is the future trend, so the system adopts C/B-S architecture [9, 10]. For learners, there are mobile phone clients and web pages, teachers use web pages, and the client and web pages are responsible for processing human-computer data interaction. The web server is developed with the thinkphp framework of MVC architecture and is responsible for voice data preprocessing, pronunciation quality evaluation, data storage, etc.
- (2) Language acquisition path
 The input and output of the language must be close, that is, interaction is required. Learning in interaction, human-to-human interaction is the best learning situation, but in reality, the lack of teacher resources makes it difficult to have one-to-one interaction Opportunity to learn. Therefore, it is necessary to use computer technology to achieve a similar “interactive” effect. Understand, the input of the language is understandable, and it is $x + 1$ of the learner’s language level. Assuming that the actual level of a language learner is x , the input information needs to be one level

higher than that of the learner [11]. If it is $x + 0$, it will not help the learner. If it is $x + 2$, it is difficult for the learner to digest, so it is difficult to continue to interact.

In layman's terms, collaboration is to give learners a lot of information elements other than verbal content, including context, language form, language style, and so on. Output means that the learner receives input and produces output through imitation. In oral learning, especially the film and television English imitating reading aloud questions, the basic input and output elements include "sight", "listening", "speaking" and "acting".

2.4 Optimization Method of MLP/HMM System Training

(1) The bottleneck layer characteristics of MLP

Generally, the posterior features output by MLP have higher dimensionality, and the difference between the values is large, so logarithm and Principal Component Analysis (PCA) dimensionality reduction processing should be performed on this vector. Using BN layer features to replace the posterior probability features of the output layer has two obvious advantages: not only can ensure reasonable feature vector dimensions without relying on dimensionality reduction, but also will not lose feature information due to dimensionality reduction. Thereby ensuring the integrity of the feature information.

(2) Segmentation of MLP training data annotated text

In the hybrid HMM/ANN speech recognition system, the training of ANN is a supervised learning process. Annotated text with detailed time segmentation corresponding to the voice data is required. Manual labeling is often time-consuming and limited in quantity, which cannot meet the requirements of large-scale training data.

(3) New model initialization method

For the training of a new model, in addition to modeling from the very first basic approach, HERest also provides another model initialization method that maps a fully trained model to a new data set. When using HERest training, the training list is listed in the form of feature pairs, the first of each pair is the training data under the corresponding original model, used to estimate the forward and backward probability, and the other is used as the new Model parameter estimation and update. Use the prefixes HPARM1 and HPARM2 in the configuration file to set the parameters of the original data and the new data respectively.

(4) MLAN system modeling on multi-accent data

When training the GMM-HMM model, it is necessary to perform the feedforward of the trained two-level MLP on the standard Mandarin data, and the process is similar. The resulting two-part MLAN features.

The MLAN features of standard Mandarin that are fed forward through the two-level MLP in the MLAN system not only include the original acoustic characteristics (PLP), but also because of the "filtering" of the second-level MLP, it has the ability to adapt to specific accent data. The training data with accent characteristics is greatly improved, and the sparsity of specific accent data is solved well.

At the same time, the final MLAN features of the accent data not only retain the acoustic features of the original PLP, but also because it has passed the first-level MLP trained by standard Mandarin data, these features “absorb” the accent to a greater extent Commonality between data and standard data.

2.5 Binomial Logistic Regression Model

Assuming that the corresponding categories of the positive and negative samples in the binary classification problem are C_1 and C_2 , given the feature vector x of the sample, the posterior probability of the sample belonging to the positive sample C_1 is defined as:

$$p(C_2|x) = y(x) = \sigma(w^T x + b) \sigma(z) = \frac{1}{1 - e^{-z}} \quad (1)$$

Among them, $\sigma()$ is a sigmoid function, and w and b are model parameters to be learned. Obviously, the conditional probability that the sample C_2 belongs to the negative sample category is:

$$p(C_2|x) = 1 - p(C_1|x) = \frac{e^{-w^T x - b}}{1 - e^{-w^T x - b}} \quad (2)$$

In the problem of phoneme pronunciation error detection, we define the samples with correct (Correct, C) pronunciation as positive samples.

3 The Design Experiment Research of English Speech Scoring Data System Based on Neural Network Algorithm

3.1 Experiment Preparation

The disadvantage of the traditional BP neural network is that the convergence speed is slow, and it is easily limited to local minimum points. For these two drawbacks, this article will increase the momentum factor, and the learning rate will be adaptive. The combination of the two makes up for the shortcomings of the traditional BP network. In order to be able to compare with the traditional BP neural network, in the experiment, the speech and recognition data, steps, and training methods are all the same as the above, and the parameters extracted by the hybrid MFCC parameter extraction method are selected as the feature parameters.

3.2 Sources of Experimental Data

The training samples used in this experiment were produced by 10 testers who pronounced the numbers 0–5, and 1000 samples were obtained. Among them, 400 training samples are obtained from the first 2 pronunciations, and 600 test samples are obtained from the last 3 pronunciations.

4 The Design Experiment Research Analysis of English Speech Scoring Data System Based on Neural Network Algorithm

4.1 Performance Comparison Before and After Neural Network Improvement

The recognition results are shown in Table 1:

Table 1. Recognition results of optimized BP algorithm

Voice signal	Number of test samples	Number of correct recognition
0	15	14
1	14	12
2	16	15
3	12	10
4	15	15
5	17	16

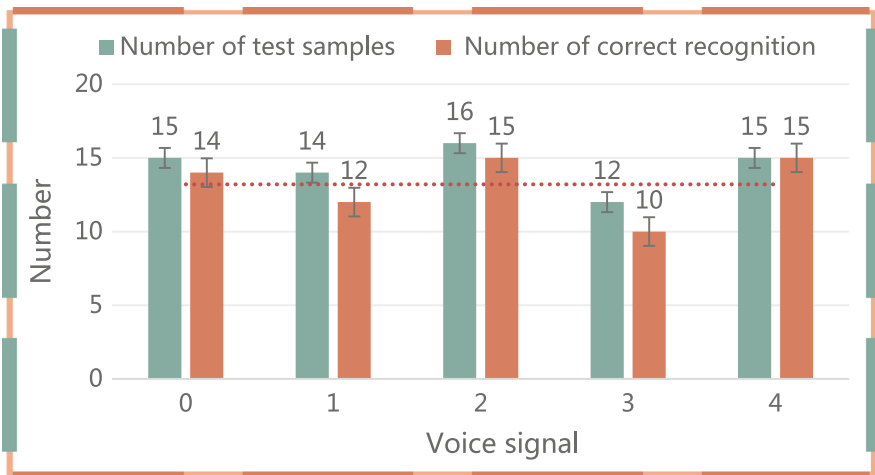


Fig. 1. Recognition result of optimized BP algorithm

As shown in Fig. 1, it can be seen that the improved BP algorithm has a higher oral English recognition rate than the traditional BP algorithm by about 10%, and the

training performance and recognition rate of the optimized algorithm is much better than the traditional algorithm.

4.2 Performance Comparison of Different Parameters

The characteristics of a piece of speech can be reflected from the characteristic parameters, so its accuracy has a great influence on the recognition rate of speech. The optimal number of hidden layer neurons selected in this article, that is, the number of neurons in the first to fifth options is set to 20, and the test samples are 250, 240, 230, 220, and 210 respectively. The final experimental results are shown in Table 2.

Table 2. Comparison of recognition performance of different parameters

Feature parameter scheme	Number of samples	Number of recognition
LPCC	250	247
MFCC	240	205
Hybrid MFCC	230	206
LPCC+MFCC	220	197
LPCC+Hybrid MFCC	210	188

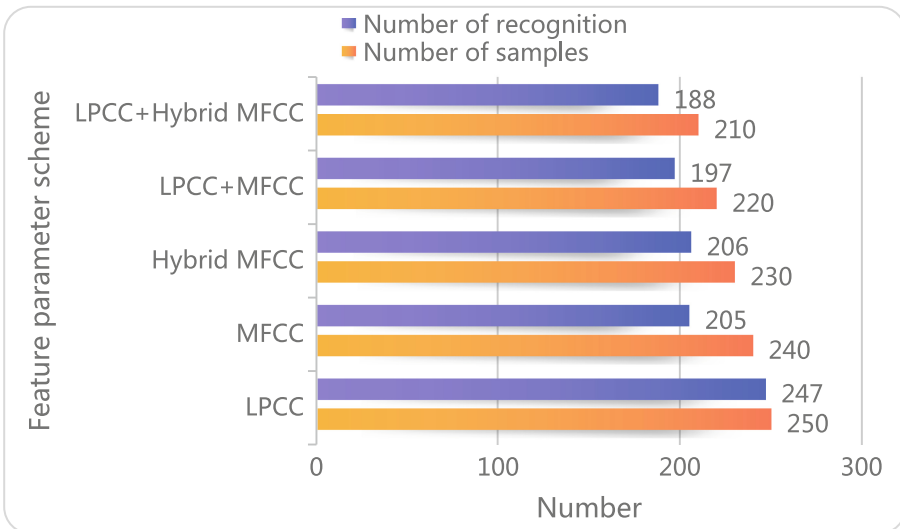


Fig. 2. Comparison of recognition performance of different parameters

As shown in Fig. 2, in the selection of parameter features, whether it is LPCC, MFCC, hybrid MFCC, MFCC and LPCC hybrid, LPCC and hybrid MFCC, the system

can accurately identify most words. Among these schemes, the recognition rates of the third, fourth, and fifth schemes are not much different, and the pattern recognition rate of the combination of LPCC and hybrid MFCC is slightly higher. Experiments have also confirmed that MFCC parameters have better performance than LPCC parameters, because MFCC is based on the auditory characteristics of the human ear, and LPCC is a synthetic parameter, mainly derived from the assumption of the all-pole model, so the recognition rate of MFCC is slightly higher. Also has good anti-noise performance. MFCC is based on the auditory characteristics of the human ear, so the logarithm of the Hz-Mel frequency shows a corresponding relationship. There are more filters in the low frequency part, and less in the middle and high frequency parts. That is to say, the higher the frequency, the lower the accuracy. Hybrid MFCC is compatible with the characteristics of the low, medium, and high frequency bands, so it is higher than MFCC in terms of recognition, which also confirms the conclusion that hybrid MFCC has better performance than traditional MFCC parameters. Finally, after mixing the two parameters of hybrid MFCC and LPCC, because it has the advantages of hybrid MFCC and LPCC at the same time, it makes the characteristic parameters more expressive and the recognition effect is more ideal.

5 Conclusions

This article mainly studies the pronunciation error detection algorithm related to the text. The algorithm is only suitable for the oral reading learning mode, but the more practical application scenario is to evaluate the user's arbitrary free speech. Spoken language evaluation that is not related to text usually requires a voice recognition system to recognize the user's voice in advance, and then use text-related strategies for processing. Therefore, its accuracy largely depends on the performance of the speech recognition system. Although deep learning greatly reduces the word error rate of speech recognition, the performance of non-standard pronunciation for beginners needs to be further improved.

References

1. Spille, C., Ewert, S.D., Kollmeier, B., et al.: Predicting speech intelligibility with deep neural networks. *Comput. Speech Lang.* **48**, 51–66 (2018)
2. Seki, H., Yamamoto, K., Akiba, T., et al.: Discriminative learning of filterbank layer within deep neural network based speech recognition for speaker adaptation. *IEICE Trans. Inf. Syst.* **E102.D(2)**, 364–374 (2019)
3. Yakoub, M.S., et al.: Improving dysarthric speech recognition using empirical mode decomposition and convolutional neural network. *EURASIP J. Audio Speech Music Process.* **2020(1)**, 1–7 (2020)
4. Kaur, G., Srivastava, M., Kumar, A.: Genetic algorithm for combined speaker and speech recognition using deep neural networks. *J. Telecommun. Inf. Technol.* **2(2)**, 23–31 (2018)
5. Pawar, M.D., Kokate, R.D.: Convolution neural network based automatic speech emotion recognition using Mel-frequency Cepstrum coefficients. *Multimedia Tools Appl.* **80(10)**, 15563–15587 (2021). <https://doi.org/10.1007/s11042-020-10329-2>

6. Min, S.K., Kim, H.S.: Compromised feature normalization method for deep neural network based speech recognition*. *Phonetics Speech Sci.* **12**(3), 65–71 (2020)
7. Vryzas, N., Vrysis, L., Masiola, M., et al.: Continuous speech emotion recognition with convolutional neural networks. *J. Audio Eng. Soc.* **68**(1/2), 14–24 (2020). Audio Engineering Society
8. Masumura, R., Asami, T., Oba, T., et al.: Latent words recurrent neural network language models for automatic speech recognition. *IEICE Trans. Inf. Syst.* **E102.D**(12), 2557–2567 (2019)
9. Zoughi, T., Homayounpour, M.M.: Adaptive windows convolutional neural network for speech recognition. *Signal Data Process.* **15**(3), 13–30 (2018)
10. Mcshane, P., Stewart, D.: Liveness verification using deep neural network based visual speech recognition. *Int. J. Multimedia Image Process.* **8**(1), 380–388 (2018)
11. Nagajyothi, D., Siddaiah, P.: Speech recognition using convolutional neural networks. *Int. J. Eng. Technol.* **7**(4), 133–137 (2018)



Financial Risk Forecast Model of Small and Medium-Sized Enterprises Based on Neural Network Algorithm

Yu Tao^(✉)

Aba Normal University, Aba, Sichuan, China

Abstract. With the continuous development of the economy and the advancement of technology, the number of enterprises in our country is increasing. With the continuous expansion of scale, the financial risks of enterprises are gradually exposed. This article focuses on the financial risk prediction of small and medium-sized enterprises based on neural network algorithms, and understands the financial risk prediction of small and medium-sized enterprises and related theories of neural network algorithms on the basis of literature data, and then analyzes the financial risk prediction model of small and medium-sized enterprises based on neural network algorithms, then carry out construction, and verify the constructed model using examples. The results of experimental verification show that the model constructed in this paper has a maximum risk prediction accuracy of 94.56% .

Keywords: Neural network algorithm · Financial risk · Risk prediction · Model construction

1 Inductions

As China's capital market is becoming more and more perfect, the fair competition among enterprises has become more and more in line with the laws of market economy, and more and more people have joined the investment of financial liabilities [1, 2]. However, opportunities and risks always coexist. The increasingly open financial market makes the early warning of corporate financial status risks more and more critical [3, 4]. At present, China is in the era of great transformation of building a well-off society, and it is also in the era of economic structural transformation [5, 6]. The big change also means huge opportunities and risks, which is particularly critical for managing the financial risks of enterprises. As the domestic securities market is experiencing unprecedented instability due to the epidemic, it is very important for companies and investors to predict corporate financial risks under such market fluctuations [7, 8].

Regarding financial risk prediction research, some researchers described the past formation, current progress and future development direction of deep learning from the perspective of the real world. Different from the opinions of the academic circles, researchers focus more on the practical application of deep learning, and point out that deep learning is very important to the practical environment, and it is also very helpful to

understand the progress and development trend of deep learning [9]. Some researchers also pointed out that there is an obvious connection between industrial environmental risks and corporate financial risks. The company's external conditions will also affect the company's financial risks. Therefore, case studies are conducted through system dynamics, and the risk indicator framework of the environmental protection industry is constructed, which involves industrial resource risks, industrial competition risks, and industrial life cycle risks, and came to the conclusion that the risks of the environmental industry must be related to the economic risks of enterprises, and system dynamics can better represent the logical connection between the two [10]. Some researchers have carried out research on early warning indicators for small and medium-sized enterprises, and introduced quality indicators such as corporate working environment, internal management, external environment, operating environment, solvability indicators, and profitability indicators for analysis, and integrated certain data for audit, research found that the introduction of quality indicators can predict financial risks well [11]. In summary, our country has a lot of research and application of corporate financial risk, but corporate financial risk is a dynamic and complex process, and the early warning models of different environments also need to be changed. This is something that static models cannot adapt. Therefore, the introduction of neural network model can solve this problem well.

This paper studies the financial risk prediction of small and medium-sized enterprises based on neural network algorithm, analyzes the causes of financial risk of small and medium-sized enterprises and the advantages of BP neural network in financial risk prediction on the basis of literature data, and then analyzes the financial risk of small and medium-sized enterprises based on neural network algorithm. The risk prediction model is constructed, and the constructed model is tested, and relevant conclusions can be drawn through the test.

2 Research on Financial Risk Prediction of SMEs

2.1 Causes of Financial Risks for SMEs

(1) The complexity of the external environment.

An enterprise has no independent business. It must be in a complex external environment and closely related to these environments, including internal and external economic environments, legal constraints, and natural conditions [12]. Most external environments are objectively irreversible. The enterprise itself only needs to adapt and is unlikely to change the environment. Therefore, for this reason, companies need to improve their quality and adaptability so that they can be internal without being restricted by the external environment.

(2) The capital structure is unbalanced.

The capital structure of an enterprise has a direct impact on its financial status. A balanced capital structure will affect solvency, generate financial risks and increase financial costs. In particular, too little capital accumulation will reduce profitability, too much capital accumulation will lead to an imbalance of capital ratios, and too much debt will significantly increase the company's financial risks.

(3) Imperfect governance structure.

The corporate governance structure is a control and balance relationship that restricts the distribution of rights and responsibilities between shareholders and directors. A good governance structure allows owners and operators to perform their duties and improve the operational performance of their businesses. On the contrary, the chaotic structure where shareholders and directors are solely responsible causes unnecessary internal friction, reduces the efficiency of the company, and causes financial problems.

(4) Poor asset liquidity.

Asset liquidity is the liquidity of company assets. However, to meet the requirements of daily operating funds, commercial assets must be reasonably realizable to meet operating needs. Insufficient asset liquidity is particularly obvious in situations such as overstocking and insufficient funds. Insufficient liquidity of company assets will have a serious impact on the company's financial status.

2.2 Advantages of BP Neural Network in Financial Risk Prediction

- (1) The operation of BP neural network is similar to the operation of students with self-learning ability. It can adapt to the environment. After many attempts to find the inner law, it can perform calculation, identification or process control. Specifically, students can absorb and learn by themselves through the knowledge imparted by teachers, and make decisions and choices based on existing knowledge of similar situations. Similarly, the BP neural network directly introduces new financial data after the artificial training data. The neural network can automatically receive the results according to the previously learned skills, so it is relatively simple and intelligent.
- (2) The comparative analysis of financial risk prediction models shows that the accuracy of artificial neural network model prediction is relatively high, and high accuracy is one of the important purposes of financial risk prediction.
- (3) The processing of artificial neural network prediction is mainly done by computer. The computer runs faster than the human brain and can perform multiple functions at the same time, which significantly speeds up the prediction task and saves time.

2.3 BP Neural Network Algorithm

Because it is habitually believed that a neuron will receive the influence of multiple other neurons, it is described as n interconnected models, and a neuron counting model is established. Neurons have certain input and output states, as shown in Fig. 1. However, the state of the neuron itself also determines whether its input and output system can be used. That is, each neuron receives the information transmitted by the other $n - 1$ neurons, thereby forming nerve excitement and impulse. As long as other conditions remain unchanged, all stimuli will work. When the total input quantity exceeds a certain limit, the possibility of action will be generated and delivered at the fastest rate. When the total number of input stimuli is less than the threshold, the neuron enters a stopped state and is not excited, and at the same time, it will not form a stimulus output for a long period of time.

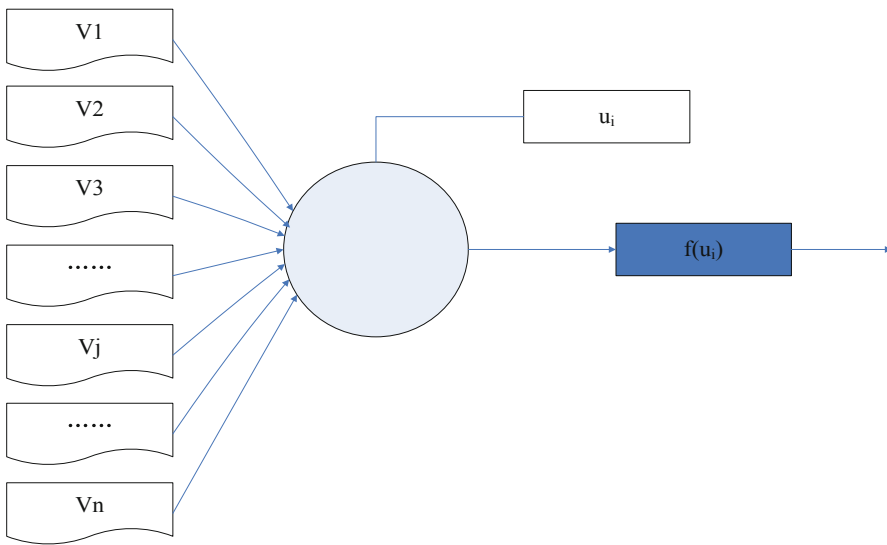


Fig. 1. BP neural network calculation model

For the i -th neuron among n interconnected neurons, the sum of external inputs affects its state, and the state of the i -th neuron is

The output is formed in the form of a certain function, namely:

$$u_i = \sum_{j=1}^i w_{ij}v_j - \theta_i \tag{1}$$

$$v_i = f(u_i) \tag{2}$$

Where:

w_{ji} ----Represents the connection strength between neuron i and neuron j (simulating the connection strength of the synapses between biological neurons), here, we call it the connection weight;

u_i ----Represents the state of neuron i ;

v_j ----Represents the output of neuron j , which is also an input of neuron i ;

θ_i ----- Represents the threshold of neuron i .

3 Construction of SME Financial Risk Prediction Model Based on Neural Network Algorithm

3.1 Predictive Indicators

Early warning indicators directly determine the rationality of the model and the effectiveness of early warning, so it is very important to choose appropriate indicators. Reasonable early warning indicators can objectively, comprehensively and dynamically reflect the

financial status of an enterprise. The complete early warning indicator system is composed of interrelated financial indicators. Changing the attribute value of one index will change the attribute value of another index. This feature makes the early warning result more objective and logical, but it also increases the early warning cost of the model. Therefore, the selection of scientific and logical early warning indicators must be consistent with the predictive ability of the model itself. The impact of early warning should also be considered.

The company's financial status perfectly reflects the company's operating conditions and operating environment. The company's business strategy has problems such as low management efficiency, investment losses, and blind expansion. Or, the company's business environment has deteriorated, such as the introduction of stronger competition strategies, policy constraints, and the implementation of regulatory systems by competitors. All of these can deteriorate the value of specific financial ratios, ultimately increase financial risks, and affect the interests of themselves or their stakeholders. Therefore, the financial risk discussed in this article is a characteristic value that can reflect and quantify the changes in the financial status of enterprises in the operating environment of China's SMEs. The high characteristic value indicates that the company's financial status is changing in the wrong direction, and it can be concluded that there is a problem with the company's business strategy or operating environment. If this characteristic value is small and within a certain range, it means that we are moving towards a better opposite change, at least there is no problem with the company's business strategy or business environment.

Based on the above considerations, and with reference to the research results of well-known scholars and the regulatory standards of China and the United States, this article provides nine early warning indicators to better reflect the company's financial status. The results are shown in Table 1:

Table 1. Early warning indicators

	Index	Relative to financial risk
Capital adequacy	A1-income ratio	Reverse indicator
	A2-Debt-to-equity ratio	Positive index
Profitability	X3-net return rate	Converse indicators
	X4-total net interest rate	Reverse indicator
Solvency	X5-current ratio	Reverse indicator
	X6-Asset-liability ratio	Positive index
Growth ability	X7-amount of payout increase and income ratio	Reverse indicator
	X8-amount of payout change rate	Reverse indicator
	X9-income growth rate	Moderate index

3.2 BP Neural Network Model

Through the nine common factors obtained above, the common factors are used as new variables in the neural network BP, and then the classification results are obtained through training and testing, and financial risk prediction research is carried out.

(1) Input vector and target vector

The input vector selection is generally based on two principles. One is that the input variables have a great influence on the output and can be output, and the other is that there is no or little correlation between the variables. Since the number of nodes in the input layer mainly depends on the dimensionality of the input data, the number of common factors obtained by dimensionality reduction through factor analysis is used as the number of input layers of the BP neural network. The selection of the output layer is relatively easy, and the determination is based on the type of output data and the size of the data required to represent the type.

The value of F is obtained through factor analysis, as the input level data of the BP neural network, so the input level has a total of nine dimensions. According to whether the company receives special treatment, it can be divided into FX companies and non-FX companies. Specifically, 0 represents a non-FX company, and 1 represents an FX company. Since the output is only 0 or 1, the output plane is defined in one dimension, and the output plane is also the target vector.

(2) Network topology

This article uses the most common 3-level BP neural network. This is also the most basic and classic topology. The input plane here is 9-dimensional, and the output plane is 1-dimensional.

Determining the number of hidden layer nodes plays an important role in the overall function of the neural network. If the number of nodes is too small, the ability of the network to output information samples is weak, and if the number of hidden layer nodes is too much, the phenomenon of “overmapping” will appear, reducing its generalization. Under normal circumstances, the number of hidden layers is determined empirically and combined with the actual results of the model. Adjust the number of neurons in the network grid layer to improve error accuracy and make training results clearer. The calculation formula for the number of hidden layers can get a certain range of hidden layer nodes, and then the number of hidden layer nodes increases continuously throughout the training process. Finally, when the number of hidden layers is 13 nodes, this is the most effective training result.

(3) Network related parameters

According to the characteristics of the data sample and the network structure, the nonlinear function of the BP neural network is integrated through the S-type transfer function, so the hidden layer is generally the S-type transfer function, the tansig function is defined as the transfer function, and the output value of the hidden layer is controlled as $(-1, +1)$. Since the output value defined in this article is 0 or 1, the logsig function is used to determine the transfer function from the hidden layer to the output layer.

This article uses an improved L-M algorithm for network training. It has the fastest convergence speed and is very suitable for small and medium-sized networks. The network training function in this article uses `trainlm`, because the batch mode `trainlm` training function is provided in the MATLAB 8.0 neural network toolbox.

4 Empirical Analysis

In order to verify the effectiveness and superiority of the model proposed in this paper, this paper uses factor analysis (FA) and BP neural network methods to compare the two models represented in the model. The model used in this article is used to train a BP neural network after dimensionality reduction through factor analysis (for convenience, the following is represented by FA-BP). Collecting three-year financial data of small companies, analyzing them with different methods, and getting the forecast results. The results are shown in Table 2:

Table 2. Model experiment comparison results

	Model 1	Model 2	Model 3
Year 1	76.10%	90.21%	90.56%
Year 2	80.89%	90.21%	95.34%
Year 3	71.32%	71.32%	90.56%

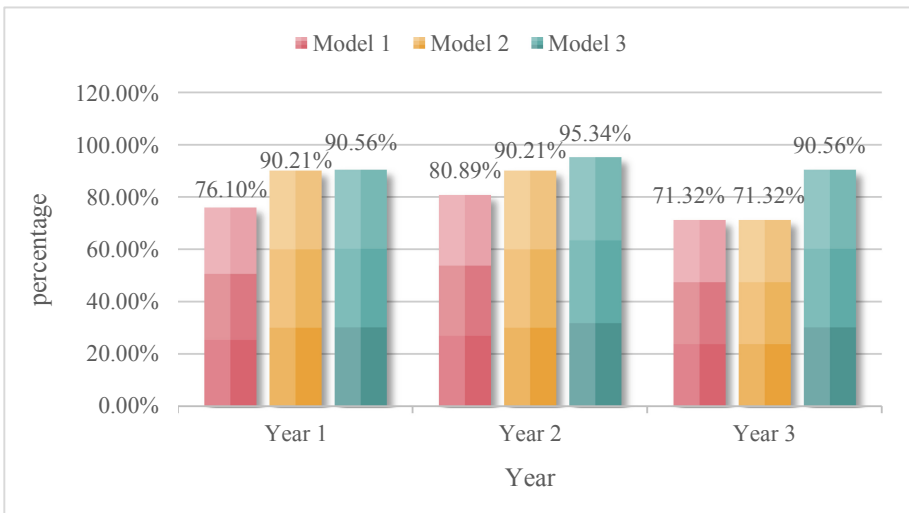


Fig. 2. Model experiment comparison results

It can be seen from Fig. 2 that among the three models, Model 3 has a three-year comprehensive judgment rate higher than the other two, reaching 90.56%, 95.34% and

90.56% respectively. Model 1 has the lowest accuracy of BP neural network, which are 76.10%, 80.89% and 71.32% respectively. Therefore, without considering other situations, the model proposed in this paper is better than the other two models and can better meet the accuracy requirements.

5 Conclusions

This article studies the financial risk prediction of small and medium-sized enterprises based on neural network algorithms. After understanding the relevant theories, constructs the financial risk prediction models of small and medium-sized enterprises based on neural network algorithms, and tests the constructed models. By introducing traditional the factor analysis method and BP neural network method are compared, and the experimental results show that the accuracy of the model constructed in this article is higher than these two models.

References

1. Song, Y., Peng, Y.: A MCDM-based evaluation approach for imbalanced classification methods in financial risk prediction. *IEEE Access* **7**, 84897–84906 (2019)
2. Ca Rdo, S.R., Barbosa-Povoa, A.P., Relvas, S.: Integrating financial risk measures into the design and planning of closed-loop supply chains. *Comput. Chem. Eng.* **85**, 105–123 (2016)
3. Ramos-Perez, E., Alonso-Gonzalez, P.J., Nunez-Velazquez, J.J.: Forecasting volatility with a stacked model based on a hybridized Artificial Neural Network. *Expert Syst. Appl.* **129**, 1–9 (2019)
4. Prades, J., Varghese, B., Reano, C., et al.: Multi-tenant virtual GPUs for optimising performance of a financial risk application. *J. Parallel Distrib. Comput.* **108**, 28–44 (2016)
5. Li, B., Arreola-Risa, A.: Financial risk, inventory decision and process improvement for a firm with random capacity. *Eur. J. Oper. Res.* **260**(1), 183–194 (2017)
6. Barnes, K., Mukherji, A., Mullen, P., et al.: Financial risk protection from social health insurance. *J. Health Econ.* **55**, 14–29 (2017)
7. Santibañez-Aguilar, J.E., Guillen-Gosálbez, G., Morales-Rodriguez, R., et al.: Financial risk assessment and optimal planning of biofuels supply chains under uncertainty. *Bioenergy Res.* **9**(4), 1–17 (2016)
8. Dellana, S., West, D.: Survival analysis of supply chain financial risk. *J. Risk Finance* **17**(2), 130–151 (2016)
9. Sadorsky, P.: Carbon price volatility and financial risk management. *J. Energy Mark.* **7**(1), 83–102 (2016)
10. Kwack, B.S., Rhee, J.H.: Review of research on corporate social responsibility of small- and medium-sized enterprises and recommendations for future research direction in Korea. *J. Strateg. Manage.* **22**(3), 21–50 (2019)
11. Valipour, M.: Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms. *Meteorol. Appl.* **23**(1), 91–100 (2016)
12. Grossberg, S.: Classical and instrumental learning by neural networks **4**(1), 1 (2019)



The Analysis of COVID-19 Propagation Process Based on Population Flow and Cellular Automaton Model

Congxiao Zhang^(✉)

College of Science, Beijing Jiaotong University, Beijing, China

Abstract. Developments in the information ages coming. The prevalence of covid-19 destroys economic and social development. This paper uses the data of COVID-19 epidemic in China from January 11, 2020 to February 9th to observe the trend of the number of confirmed cases and analyze the spatiotemporal distribution characteristics of the patients. On this basis, we establish the CA model and focus on the influence of different population activities on the transmission of infectious diseases by changing the parameters. Finally, it is concluded that inter provincial migration does have a certain impact on the spread of the epidemic, but its impact capacity is related to the geographical location of the target province; for individuals, the stronger the activity capacity is, the faster it spreads.

Keywords: CA model · COVID-19 · Epidemiological statistics · Population dynamics · Informatization

1 Introduction

At the end of 2019, COVID-19 appeared in Wuhan, China, and then spread rapidly across the country. Most of the first batch of cases came from a seafood market. The relevant health management departments in Wuhan closed the seafood market on January 1, 2020, began to isolate patients on January 7, and released the relevant information of the virus genome sequence on January 10. Therefore, the global public health department could take certain screening and disease prevention measures, and gradually determined the diagnosis process. The initial stage of transmission of this virus is similar to that of SARS coronavirus in 2003. So in the process of research, the relevant information of SARS is referred to.

G.Ch. Sirakoulis and I. Karafyllidis study the effect of two population movement parameters on the epidemic propagation: the distance of movement and the percentage of the population moves [1, 2]. Tzai-Hung Wen, Min-Hau Lin and Chi-Tai Fang investigated the effects of local environmental and demographic characteristics to clarify the role of the daily commute in Dengue transmission. They also analysed the clustering pattern of space-time distances between commuting and noncommuting Dengue cases from June 2007 to January 2008 in Tainan city, Taiwan [3, 4]. Jorge A. Falcon-Lezama, Ruth A. Martinez-Vega focus on the relationship between Dengue transmission and People's

day-to-day movement. They use a simple model to examine the fundamental roles of broad demographic and spatial structures in epidemic initiation, growth and control [5]. Ping Zhang and Peter M. Atkinson selected the main factors of the urbanization process (urban sprawl, land use development, population movement and age structure) to model the impact of urbanization on infectious disease transmission by integrating a CA land use development model and CA epidemic model in S-Plus [6]. Jayanthi Rajarethinam, Janet Ong and Shi-Hui Lim try to understand the implication of human movement on disease spread. A retrospective study was conducted using aggregated and anonymized mobile phone data to examine movement from the cluster to identify areas of possible transmission. Their result supports the proof of concept of using mobile phone data to approximate population movement, thus identifying areas at risk of disease transmission [7].

Refer to the above literature, we can simplify the transmission process of COVID-19 can be as follows: the infected person spreads the virus to the susceptible population in the incubation period, making it become a new infected person. After the incubation period, due to the characteristics of the virus, the transmissibility after the onset of the disease is greatly reduced [8].

The range of daily activities is a key factor which affects the spread of disease. In addition, during the Spring Festival, China's population flow is very frequent, and the percentage of population emigration will also become a key factor affecting the spread of the disease.

2 Data Acquisition, Cleaning and CA Modeling

2.1 Data

The date of selected data is from January 11 to February 9, 2020.

The data describe the spread of COVID-19, including the cumulative number of confirmed cases, newly diagnosed cases, deaths and cures, are from the official website of the National Health Council of China. The relevant data of population migration about the percentage of Wuhan's daily population moving out to each province are obtained from Baidu by Python.

2.2 Statistical Analysis Method

2.2.1 Basic Trend

The cumulative number of confirmed cases in each province and the percentage of population migration from Wuhan to the province are shown as line chart.

2.2.2 Analysis of Temporal and Spatial Characteristics

Plot the spatial distribution of the cumulative number of confirmed cases and the percentage of population migration from Wuhan to other provinces.

2.2.3 The Impact of Population Activity on Disease Transmission: CA Model

CA Model is for physical system with discrete space and time, which is used to describe local interaction. It is widely used in the field of complex systems. In this paper, we use CA cellular machine to simulate the spread of covid-19 under different population activity intensity by setting different parameters [9].

CA Model is composed of n regular and uniform units. A physical quantity is used to represent the global state of CA, and the value of the physical quantity corresponding to each cell is its local state. Each cell only interacts with its neighborhood, that is, it only interacts with its neighboring cells. The state of each cell is updated simultaneously based on its neighborhood and the state of the previous time.

There are five elements in CA Model:

- ① It is composed of n units;
- ② CA Model dimension parameter w , where w_{ij} is the width of the cell in the $i - th$ row and $j - th$ column;
- ③ The neighborhood width is d ;
- ④ The state of CA Model;
- ⑤ Operation principle of CA Model [2].

According to the above principles, we construct a CA Model:

Let the CA Model be C , $C_{i,j}^t$ represents the state of unit (i, j) at time t . For the convenience of representation, we call it $C_{i,j}^t = \{P_{i,j}^t, INF_{i,j}^t, IMF_{i,j}^t\}$.

$INF_{i,j}^t$ indicates whether the unit (i, j) at time t is infected. $INF_{i,j}^t = 0$ means no infection, and $INF_{i,j}^t = 1$ indicates infection.

$P_{i,j}^t = \frac{S_{i,j}^t}{T_{i,j}^t}$, $S_{i,j}^t$ is the number of infected population in unit (i, j) at time t , $T_{i,j}^t$ represents the total number of people in unit (i, j) and $INF_{i,j}^t = 0, P_{i,j}^t = 0$.

$IMF_{i,j}^t$ indicates whether the population in unit (i, j) is immune to diseases at time t , and $IMF_{i,j}^t = 0$ is not immune; $IMF_{i,j}^t = 1$ is immune. The time of invalidation was t_{im} .

So we can draw a conclusion:

$$P_{i,j}^{t+1} = P_{i,j}^t + k \left(P_{i-1,j}^t, P_{i,j-1}^t, P_{i,j+1}^t, P_{i+1,j}^t \right) + l \left(P_{i-1,j}^t, P_{i,j-1}^t, P_{i,j+1}^t, P_{i+1,j}^t \right), d = 1$$

Where k and l represent infection coefficient respectively, $k > l$.

2.2.4 Correlation Analysis

(1) Grouping:

According to the scatter plot of population migration percentage, the provinces (municipalities) in China were divided into four levels according to the color (representing the degree of infection): (I) Henan Province and Hunan Province; (II) Chongqing City, Jiangxi Province and Guangdong Province; (III) Shaanxi Province, Guizhou Province, Anhui Province, Jiangsu Province and Zhejiang Province. (4) Other provinces (municipalities, autonomous regions, etc.). Since the percentage of population migration and infection in the fourth group were very low (almost zero) from January 11 to February 9, only the first three groups were selected as the research objects [10].

On this basis, further segmentation was conducted to study whether the impact of population migration percentage at the same level on epidemic transmission was different under different geographical conditions, On this basis, two regions with large latitude differences in each group were further selected for comparative analysis. The results showed that (I) Henan Province (North), Hunan Province (South) (II) Chongqing City (North), Guangdong Province (South) (III) Shaanxi Province (North), Guizhou Province (South).

(2) Correlation Analysis

① Covariance:

Let X be the percentage of population migration and y be the number of new confirmed cases.

If the total correlation coefficient is expressed by ρ , then there is

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{DXDY}} = \frac{E[(x - \mu_X)(Y - \mu_Y)]}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Where σ_X^2, σ_Y^2 is the total variance of sample X and sample y respectively, $\text{cov}(X, Y)$ is the covariance of X and y.

According to Cauchy-Schwarz inequality:

$$[\text{cov}(X, Y)]^2 \leq \sigma_X^2 \sigma_Y^2$$

The deformation is $\rho_{XY} \in [-1, 1]$, where the closer the linear correlation between the two variables, $|\rho_{XY}|$ is to 1.; conversely, the lower the linear correlation between the two variables, $|\rho_{XY}|$ is to 0.

② Based on (I), the sample correlation matrix is obtained.

$$\begin{pmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \cdots & r_{pp} \end{pmatrix}$$

Where $(r_{ij})_{p \times p}$ is the simple correlation coefficient of two variables:

$$r_{ij} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

③ The scatter plot of [X,Y] was drawn, and the impact of population migration percentage on the spread of the epidemic was analyzed by the curves of newly diagnosed number and the curve of population migration percentage.

3 Data Analysis

3.1 Overall Trend Analysis of Epidemic Data

Compare relevant data, We can find that: (1) compared with other provinces, Hubei Province, as the source of the epidemic, has a much higher increase in the cumulative number of confirmed cases than other provinces; (2) except for Hubei Province, as of February 9, 2020, the cumulative number of confirmed cases has not exceeded 1200; (3) except for Hubei Province, according to the number of confirmed cases, all regions can be roughly divided into severe epidemic (600–1200) There are three levels: medium (200–400) and slight (<200). The difference between the first level and the second level is more obvious. (4) In the follow-up analysis, we take January 27 as the research node, and assume that the first batch of local infected people begin to get sick on January 27, and those people are from the immigrants.

3.2 Analysis of Space-Time Cluster

3.2.1 Severe Infection Group: Henan Province (North), Hunan Province (South)

① Henan Province
Correlation Matrix

	Y	X
Y	1.00000	-0.42395
X	-0.42395	1.00000

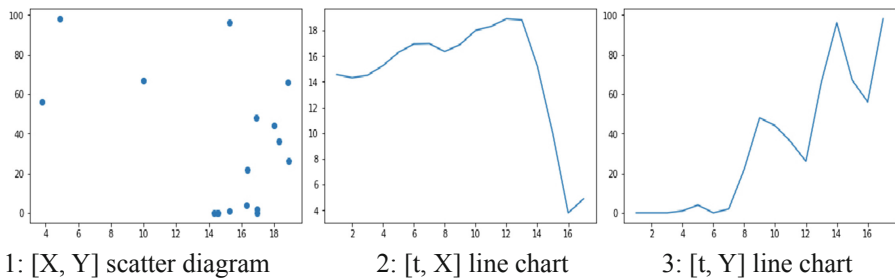
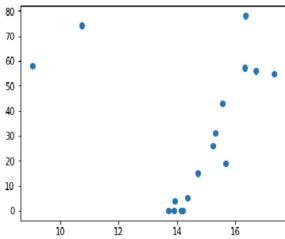


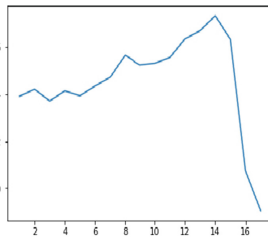
Fig. 1. Correlation test in Henan province

② Hunan Province
Correlation matrix

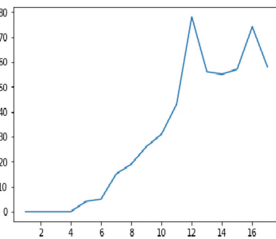
	Y	X
Y	1.000000	0.024866
X	0.024866	1.000000



1: [X, Y] scatter diagram



2: [t, X] line chart



3: [t, Y] line chart

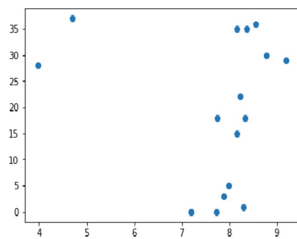
Fig. 2. Correlation test in Hunan province

3.2.2 Moderate Infection Group: Chongqing (North), Guangdong (South)

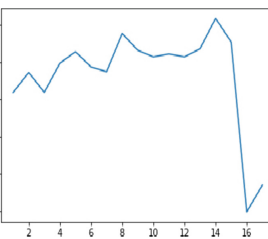
① Chongqing

Correlation Matrix

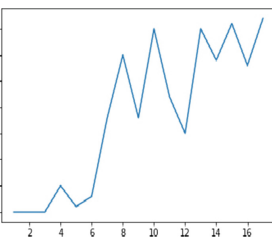
	Y	X
Y	1.000000	-0.090849
X	-0.090849	1.000000



1: [X, Y] scatter diagram



2: [t, X] line chart



3: [t, Y] line chart

Fig. 3. Correlation test in Chongqing

② Guangdong Province

Correlation Matrix

	Y	X
Y	1.000000	0.662102
X	0.662102	1.000000

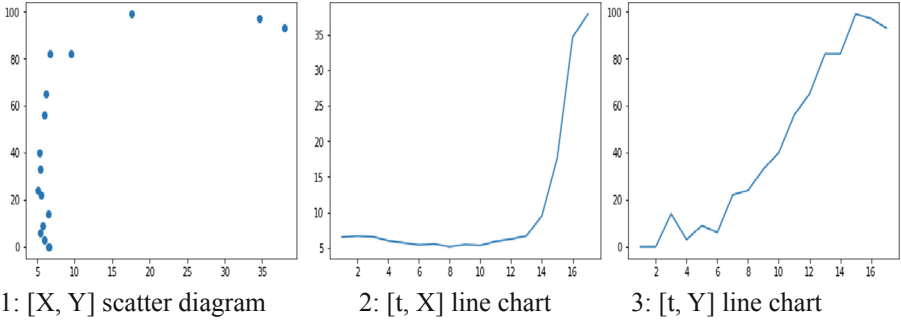


Fig. 4. Correlation test in Guangdong province

3.2.3 Mild Group: Shaanxi Province (North), Guizhou Province (South)

① Shaanxi Province

Correlation matrix

	Y	X
Y	1.00000	0.17933
X	0.17933	1.00000

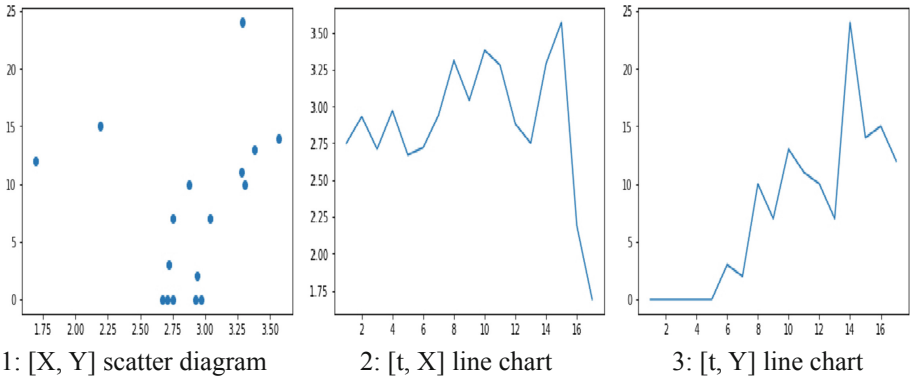


Fig. 5. Correlation test in Shaanxi province

According to Figs. 1, 2, 3, 4, 5 and 6 and the correlation coefficient matrix, it is concluded that the percentage of population migration and the number of newly diagnosed cases do not meet the linear correlation, but it does affect the spread of the epidemic.

② Guizhou Province
Correlation Matrix

	Y	X
Y	1.00000	-0.30143
X	-0.30143	1.00000

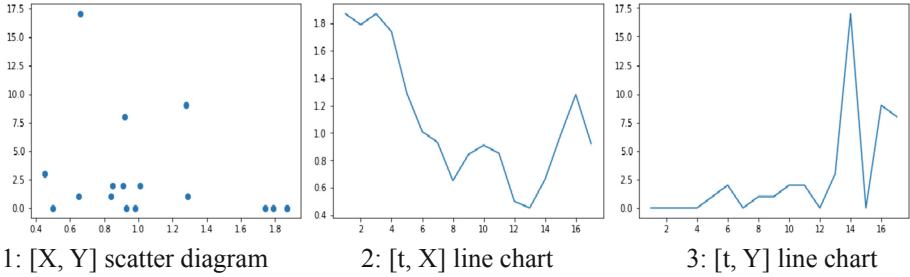


Fig. 6. Correlation test in Guizhou province

3.3 The Propagation Process of COVID-19 Simulated by CA Model

Special note: since the data are from the early stage of the epidemic, we assume that $IMF_{i,j}^t = 0$. The figure below shows the result of 220 iterations.

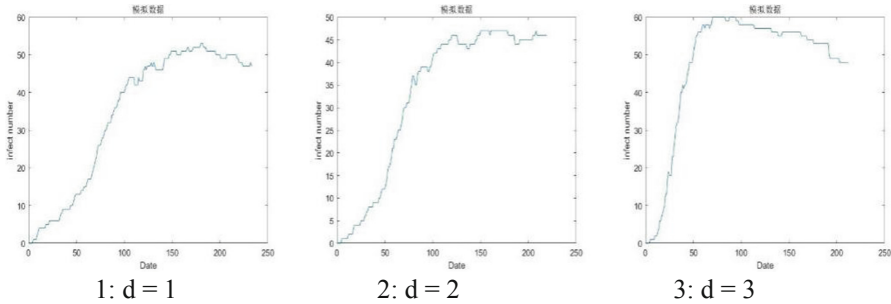


Fig. 7. The transmission process of COVID-19 simulated by CA Model

According to Fig. 7, in the natural state (without any protective measures) ① The spread of COVID-19 is faster and faster in the early stage; when the number of infected persons is close to the total number, the growth rate obviously slows down. This trend is independent of ‘d’ of the simulated population’s activity radius. ② The higher the d values, the shorter the time is needed for the early growth stage and the steeper the curve is. That is, a larger range of human activities will make it easier for the epidemic to spread. From this point of view, it is necessary to restrict the crowd activities.

4 Conclusion

According to Figs. 1, 2, 3, 4, 5, 6 and 7, the percentage of population flow mainly determines the time when the non-source area of the epidemic enters the stage of rapid

spread of the disease, but has no obvious impact on the development of the follow-up epidemic [11]. In the process of simulation of epidemic spread by CA Model, we set different d , which is the radius of population activity. We found that the larger the distance of population activity will made the epidemic spread more easily. Therefore, restricting the flow of population between provinces can reduce the spread of the epidemic, while restricting the scope of local population activities can slow down the growth of the number of infected people. Since there was no cure in the early stage of the epidemic, the number of cured people was ignored in this study. If we add follow-up data, we need to add elements on this basis to simulate the healer and change the infection probability after taking isolation and prevention measures. On the other hand, the migration mode of population can be divided into groups according to different methods of transportation by using GIS system so as to make CA Model more accurate.

References

1. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. . Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N. Engl. J. Med.* **382**(13) (2020)
2. Sirakoulis, G.Ch., Karafyllidis, I., Thanailakis, A., et al.: A cellular automaton model for the effects of population movement and vaccination on epidemic propagation. *Ecol. Model.* **133**, 209–223 (2000)
3. Tindale, L.C., et al.: Transmission interval estimates suggest pre-symptomatic spread of COVID-19 (2020)
4. Wen, T.-H., Lin, M.-H., Fang, C.-T.: Population movement and vector-borne disease transmission: differentiating spatial-temporal diffusion patterns of commuting and noncommuting Dengue cases. *Ann. Assoc. Am. Geogr.* **102**(5), 1026–1037 (2012)
5. Rajarethinam, J., et al.: Using human movement data to identify potential areas of Zika transmission: case study of the largest Zika cluster in Singapore. *Int. J. Environ. Res. Public Health* **16**, 808 (2019)
6. Calvetti, D., Hoover, A.P., Rose, J., Somersalo, E.: Metapopulation network models for understanding, predicting, and managing the coronavirus disease COVID-19. *Front. Phys.* (2020)
7. Calvetti, D., Hoover, A., Rose, J., Somersalo, E.: Bayesian dynamical estimation of the parameters of an SE(A)IR COVID-19 spread model. arXiv preprint [arXiv:2005.04365](https://arxiv.org/abs/2005.04365) (2020)
8. Shaman, J.: An estimation of undetected COVID cases. *Nature* **590**, 38–39 (2020). <https://doi.org/10.1038/d41586-020-03513-9>
9. He, X., et al.: Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020)
10. Falcon-Lezama, J.A., Martinez-Vega, R.A., Kuri-Morales, P.A., Ramos-Castaneda, J., Adams, B.: Day-to-day population movement and the management of dengue epidemics. *Soc. Math. Biol.* **78**, 2011–2033 (2016). <https://doi.org/10.1007/s11538-016-0209-6>
11. Zhang, P., Atkinson, P.M.: Modelling the effect of urbanization on the transmission of an infectious disease. *Math. Biosci.* **211**, 166–185 (2008)



Facial Image Restoration Algorithm of Dazu Rock Carvings Based on Machine Vision

Zhuying Ran^(✉)

College of Environmental Arts, Chongqing College of Architecture and Technology, Chongqing, China

Abstract. Image restoration is a technology that uses known information to supplement the content of defects, and it is one of the main research topics in the field of image processing. This paper studies the Dazu Rock Carving Face Image Restoration Algorithm based on machine vision, understands the related theories of the image restoration algorithm on the basis of literature data, and then designs the Dazu Rock Carving face image restoration based on machine vision and references it. The image restoration algorithm of this paper is used for experiments. Through experiments, it is concluded that the time-consuming and PSNR value of the improved algorithm in this paper is due to the traditional algorithm, which shows the superiority of the algorithm in this paper.

Keywords: Machine vision · Image restoration · Stone carving face · Restoration algorithm

1 Introduction

Vision is the most important way to recognize human information. Daily life, social communication, work and research always convey visual information [1, 2]. As the basis of vision, image is the main means to perceive external things and convey effective information [3, 4]. In a society with highly developed information technology, images play an increasingly important role in information dissemination. The efficient processing of image information ensures the accuracy and stability of information transmission. Image information processing is a science that studies human creation and collection, extraction and analysis, as well as the representation and use of image signals. It is a subject with intersecting characteristics. It is closely related to peripheral industries such as pattern recognition and artificial intelligence. It is used in the protection of cultural assets, Biomedical engineering, communication engineering, military public security, culture and art, etc. It has wide range of uses [5, 6]. Therefore, the research and development of image processing technology has important theoretical and practical significance for scientific progress and social practice activities [7, 8].

For the research of image restoration, some researchers have proposed a new edge detection method based on the machine vision thread detection system. It solves the problem of thread detection on the Internet through particle conversion noise reduction and local maximization processing [9]. There is also an image restoration algorithm

proposed by researchers. The algorithm uses reliability and image boundary information to define a priority function, prioritizes reliable pixel blocks, and includes structural information, so that the repair result is closer to the original image and the image, and the structure contour is more connect. However, the Criminisi algorithm also has some defects when processing images [10]. The following is an example: (1) The quality of repair largely depends on the size of the pixel block. The pixel block is too small, the calculation is time-consuming, the repair result is chaotic, and it is prone to blockage. The repair sequence is disordered and the repair result is wrong. (2) The filling order of pixel blocks also has an important influence on the repair effect. In the middle and final stages of recovery, the traditional priority algorithm will have errors, which will disrupt the recovery order and make the recovery results wrong [11]. In summary, there are many research results on image restoration technology, and there are still some problems in the image restoration algorithm that need to be resolved.

This paper studies the Dazu Rock Carving Face Image Restoration Algorithm based on machine vision, and analyzes the application fields of image restoration and the theory of the image restoration algorithm based on the literature, and then conducts the Dazu Rock Carving face image restoration algorithm based on machine vision. Experiment and verify the effectiveness of the algorithm through experimental results.

2 Research on Image Restoration Algorithms

2.1 Application Areas of Image Restoration

- (1) The main purpose of restoring the original era of digital imaging technology is to preserve precious ancient artifacts, calligraphy and painting. Many ancient artifacts, paintings and calligraphy have been damaged or scratched to varying degrees during the long-term storage and transportation process, but because such paintings and calligraphy works are often art treasures with high humanistic significance and aesthetic value, or for historical images with important historical material significance, digital image inpainting technology is the best choice [12].
- (2) Repair images and videos damaged due to transmission loss. When using a lossless transmission protocol to transmit video, music or images over the Internet, factors such as broadband will cause frequent data loss, resulting in the loss of some encoding blocks. This phenomenon makes it impossible to guarantee the surface quality of the download cost, in which case digital image restoration technology can be used.
- (3) In these damaged audio, video or image videos or old movies, peel off stains and scratches. The aesthetic taste of video viewers is often determined by the clarity or comfort of the video or picture content. Such as projected object scratches or unpleasant stains, etc., can be eliminated using digital image correction technology to restore the original vision.
- (4) Remove unnecessary parts in the image to make the image beautiful. During the shooting, due to the complexity of the background, unnecessary parts of the shot were collected. Digital image restoration technology can remove or modify these objects to prevent the observer from discovering that they have changed and make the image look comfortable and natural.

2.2 Image Restoration Algorithm

(1) Priority calculation

The priority of calculation is to determine the order of image correction, the purpose is to widen the linear contour of the image structure, so that the boundary information can be connected first.

(2) Quick search and match

After calculating the priority, first find the block with the highest priority value for repair, and search for the most similar matching block in the source area.

(3) Reliability update

When the high-priority block is restored and filled, the reliability and time of these pixels will also change, so the unknown area in the block to be restored becomes a known block, and the unknown pixel also becomes a known pixel, so need to refresh here.

2.3 Problems in Criminisi Algorithm

(1) In the priority calculation, the confidence interval is greatly affected by the pixel value of the pixel itself. The confidence term is considered to measure the amount of reliable information around the pixel, so the size of the pixel value cannot be determined. The confidence item is not safe, because the amount of information that can be trusted only describes the brightness of the known area around the pixel.

(2) In the process of multi-image restoration, the priority is calculated by multiplying the trust item and the data item, so the trust item and the data item are reduced sharply. When a part of it is reduced, the priority is too low to be accurate. Indicates the information of the other party. In this example, this problem has a significant impact on the repair effect, especially when repairing small images.

3 Facial Image Restoration of Dazu Rock Carvings Based on Machine Vision

3.1 Repair Evaluation

(1) Subjective evaluation method

Subjective evaluation is highly uncertain and subjective, and can be divided into absolute evaluation and relative evaluation. The absolute score is based on the most intuitive visual perception of the scored image. Relative evaluation is based on the subjective perception of comparing images, so the image quality is evaluated, and the evaluation result is expressed as the average score of the rater. Subjective evaluation of image quality can intuitively derive the results of image processing, and it is easy to set standards. However, subjective participants are affected by factors such as cognitive background, scoring environment, and scoring motivation, and their accuracy evaluation is easy and not convincing enough. This method requires more participants, which is time-consuming and expensive. Therefore, in practice, the method of objectively evaluating image quality is more common.

(2) Objective evaluation method

Objective image quality evaluation is a measure of the pixel error between the processed image and the original image. The method is simple in calculation, fast in calculation, easy to establish a quantitative model, and high in stability. It is often used as a template for evaluating image quality. The evaluation of image quality mainly includes average value, standard deviation, information entropy, mean square error (MSE) and signal-to-noise ratio (PSNR).

3.2 Image Restoration

This article is based on the idea of “creative” restoration strategy, because the restoration of Dazu Rock Carved Face Images is often not enough to find similar semantically similar images one by one. In the selection of restoration methods, the “original” method of copying a similar image is rejected, and the “borrowing” method is directly adopted multiple similar images. In the process of processing multiple similar images, the method of re-sampling from multiple similar images is abandoned, the target image is constructed, and new image samples are created based on the similar images to avoid the possibility of similarity and avoid high time costs.

Therefore, this paper has created a sample library, a new multi-sample restoration method based on Dazu Rock Carvings Face Image Samples, based on the Criminity algorithm and image recognition method, to address the problem of restoring Dazu Rock Carvings face images. The following describes how and how to modify the face image of Dazu Rock Carvings based on image fusion.

- (1) Selection of similar images. Using image recognition, first use the deep convergence neural network to learn the characteristics of the database sample, and then use the cosine distance measurement to select the first 10 images closest to the target image as the reference image.
- (2) Create an image sample. Use the selected reference image as the geometric reference of the image to be repaired, perform spatial correction and alignment, and then merge each pair of paired reference images to create a sample image.
- (3) Image restoration. Using the MSRoVGG method and cosine distance measurement, the generated sample image and the repaired image are checked for similarity, and they are highly similar. The sample image (less) is selected as the corresponding sample, and the sample image is restored through the block search field.

3.3 Implementation of Deep Convolutional Neural Network

The convolutional neural network in this section is implemented by the theano library in Python. There are 6 layers in total, except for the input layer, each layer contains trainable parameters. As shown in Fig. 1.

The input layer is a normalized initial segmented text image; the convergence layer C1 is composed of 22 feature maps, each feature map is composed of 30×30 neurons, and the input of each neuron is the region of the original image.

The down-sampling layer S1 uses the principle of partial image correlation for image sampling. It also contains 22 feature maps. Each feature map is composed of 15×15

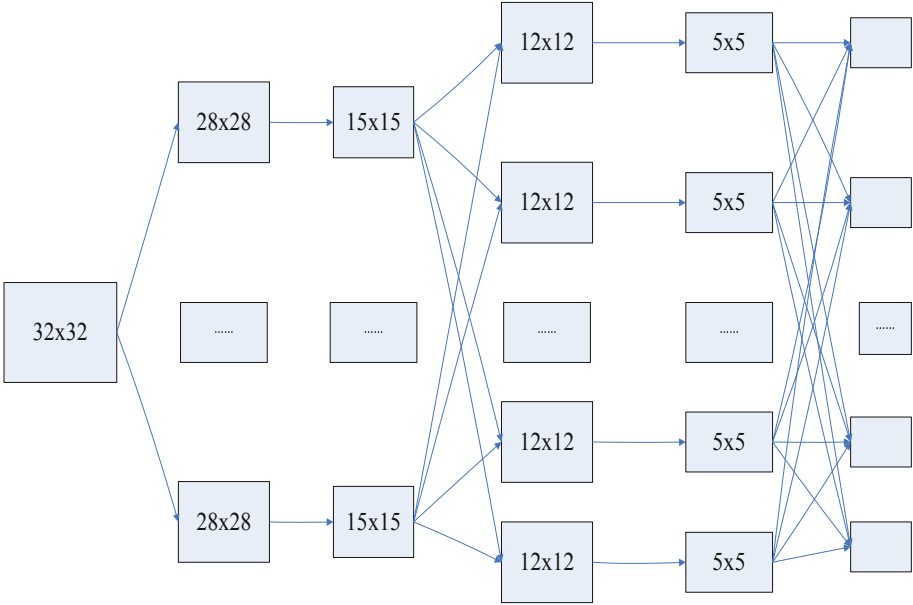


Fig. 1. Deep convolutional neural network structure

neurons, and each neuron is composed of each C1 level feature, and outputs the maximum value near 3×3 .

C2 is also a convergent plane, consisting of 51 feature maps, and each feature map is composed of 12×12 neurons.

The sampling map S2 at the bottom is composed of 51 feature maps, and the operation mode is the same as that of S1.

F is a fully connected layer consisting of 410 neurons, and each neuron is fully connected to S2. In other words, the input of each neuron is the area of all 50 feature maps in the previous layer.

L is the output level, accepts 410 inputs and produces 40 outputs.

3.4 Improvement of Repair Algorithm

- (1) The Criminisi algorithm expands the initial image information of the area to be repaired according to the edge information of the area to be repaired to meet the requirements of human optical connection. Priority determines the order of repair and affects the progress of the repair cycle.

The data element $D(p)$ has special cases, such as I_p and n . There is a vertical phenomenon, which causes the priority to be zero, which has a greater impact on the priority calculation. The reason is that special circumstances make the confidence level $C(p)$ meaningless in terms of priority. Therefore, the normalization function is introduced to smooth the data element $D(p)$, which can resist the interference of special circumstances.

(2) Improve the search and completion of the optimal template

Finding and perfecting the best template is to perfect the complete site information of the area to be repaired. The higher the accuracy, the smaller the pattern error required, and the higher the quality of the repair. This step lays the foundation for the next cycle step and plays an important role in the entire Criminisi algorithm.

This article will introduce the concept of mean value and mean square error to measure the fluctuation of the mean value and gray level of the pattern, assuming that the size of the pattern is a constant value. The size of the template is determined according to the specific characteristics of the area to be repaired in the image. If the image structure and texture information are more complex, use a small format template. If the image structure and texture information are relatively simple, a large format template will be selected. In this article, the template is set to be square, with sizes ranging from 3×3 to 9×9 . After obtaining the best template step, select the template size. The selection method is as follows.

$$E(P, (p, n)) = \frac{1}{n^2} \sum_{k=1}^{n^2} g(k) \tag{1}$$

$$D(P, (p, n)) = \frac{1}{n^2} \sum_{k=1}^{n^2} (g(k) - E(P, (p, n)))^2 \tag{2}$$

In formula (1), $E(P(p,n))$ represents the average gray value of a square pattern with side lengths n and p as the center point, where $g(k)$ represents the returned active gray value. Template the active pixel in is the pixel at the intersection of the template and the complete area, n^2 is the size of the template, in formula (2), $D(P(p,n))$ represents an edge, the length is n , and p is the center, The gray gradient changes in a square dot pattern.

4 Improve Algorithm Testing

4.1 Data Source

The image data in this article are mainly color RGB digital images of Buddhas in six different caves in three mountains. Initially, 1,200 digital images were taken. After providing some data from the Dazu Rock Carvings Research Institute, a tour was conducted. In order to increase the number of sample libraries, thousands of images were converted horizontally to obtain the same number of mirror images.

4.2 Experimental Design

This paper compares the traditional algorithm with the algorithm in this paper, and illustrates the feasibility of the algorithm in terms of the time-consuming of different repair methods and the PSNR statistical table. The experimental results are shown in Table 1:

Table 1. Comparison result of traditional algorithm and this algorithm

	T/s	PSNR/dB
Classic Criminisi algorithm	708.01	38.5
Algorithm 1	750.05	39.1
Algorithm 2	679.07	41.2
Improved algorithm in this paper	657.04	42.5

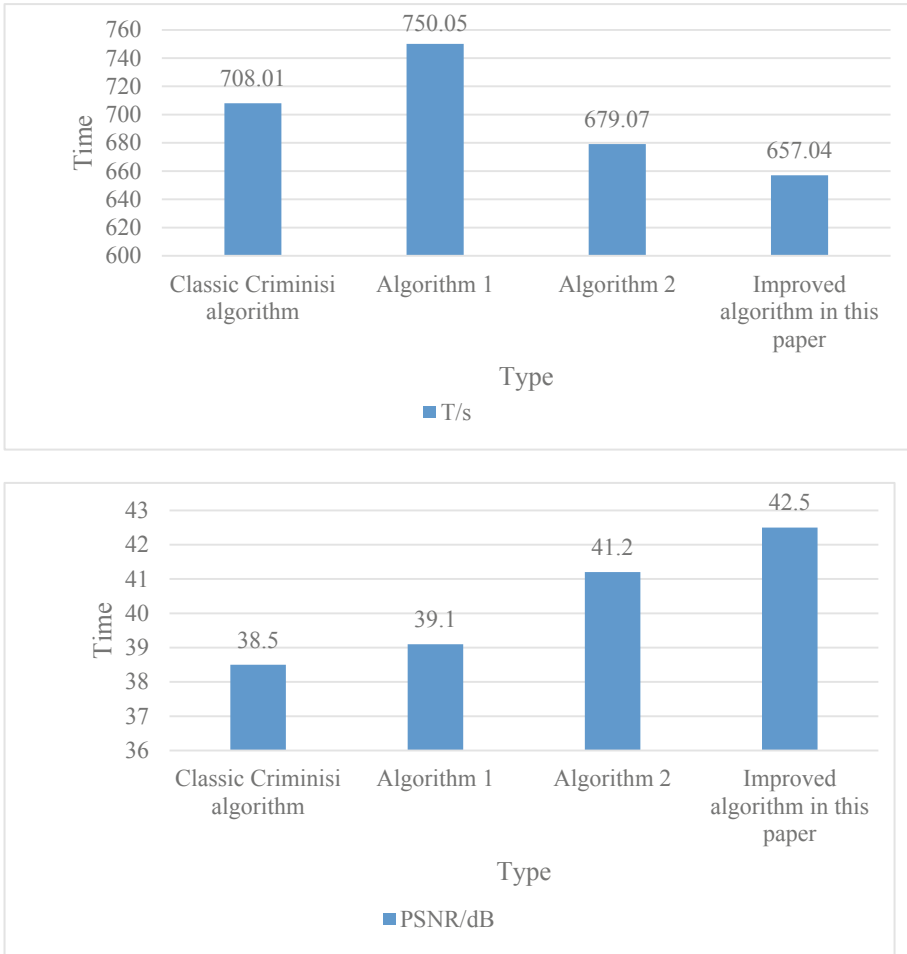


Fig. 2. Comparison result of traditional algorithm and this algorithm

It can be seen from Fig. 2 that the error information of the algorithm in this paper is less accumulated, and the average time-consuming time of the algorithm in this paper

is 650 s, which improves the repair effect while reducing time consumption, which has good practical value.

5 Conclusions

This paper studies the Dazu Rock Carving Face Image Restoration Algorithm of Machine Vision, understands related theories on the basis of literature, and then designs the Dazu Rock Carving face image restoration algorithm based on machine vision, and detects the designed algorithm. The detection result shows that the average time-consuming time of the algorithm in this paper is 650 s, which is shorter than the traditional algorithm.

References

1. Han, T.: Design and application of multicolor image identification in soil pollution component detection. *Arab. J. Geosci.* **13**(18), 1–9 (2020). <https://doi.org/10.1007/s12517-020-05873-y>
2. Chui, T.K., Tan, J., Li, Y., et al.: Validating an automated image identification process of a passive image-assisted dietary assessment method: proof of concept. *Public Health Nutr.* **23**(15), 2700–2710 (2020)
3. Zhang, R., Tian, Y., Zhang, J., et al.: Metric learning for image-based flower cultivars identification. *Plant Methods* **17**(1), 1–14 (2021)
4. Gou, B., Cheng, Y., Zhao, M., et al.: Multi-stage star image identification method of three field-of-view star sensor. *Xibei Gongye Daxue Xuebao/J. Northwestern Polytechnical Univ.* **37**(3), 541–546 (2019)
5. Garhwal, A.S., Yan, W.Q.: BIIIA: a bioinformatics-inspired image identification approach. *Multimedia Tools Appl.* **78**(8), 9537–9552 (2018). <https://doi.org/10.1007/s11042-018-6551-y>
6. Jamal, A.: Representation learning for dental image identification. *Int. J. Adv. Trends Comput. Sci. Eng.* **8**(6), 3476–3480 (2019)
7. Kasinathan, T., Uyyala, S.R.: Machine learning ensemble with image processing for pest identification and classification in field crops. *Neural Comput. Appl.* **33**(13), 7491–7504 (2021). <https://doi.org/10.1007/s00521-020-05497-z>
8. Sun, C., Wang, L., Wang, N., et al.: Image recognition technology in texture identification of marine sediment sonar image. *Complexity* **2021**(2), 1–8 (2021)
9. Kang, S., Park, H., Park, J.I.: Identification of multiple image steganographic methods using hierarchical ResNets. *IEICE Trans. Inf. Syst.* **E104.D**(2), 350–353 (2021)
10. Yi, K.Y., Seo, K.: Image based identification and localizing of drones. *Trans. Korean Inst. Electr. Eng.* **69**(2), 331–336 (2020)
11. Lysakov, K.F., Oblaukhov, K.K., Shadrin, M.Y.: Implementation of FPGA algorithms for identification of image distortion due to compression. *Optoelectron. Instrum. Data Process.* **56**(1), 28–32 (2020). <https://doi.org/10.3103/S8756699020010045>
12. Wu, S., Wang, B., Zhao, J., et al.: Virtual sample generation and ensemble learning based image source identification with small training samples. *Int. J. Digital Crime Forensics* **13**(3), 34–46 (2021)



Data Mining Model of English Translation Teaching Based on Virtual Reality Technology

Na Li^(✉)

Radio and TV University of Jilin City, Jilin 132000, Jilin, China

Abstract. In modern society, the development of information technology is becoming more and more rapid, especially virtual reality technology and data mining, which have become one of the hot spots of the current era. From the current point of view, computer technology has become an important application tool in the field of English translation. However, in the translation process, the traditional English teaching model can no longer meet the needs of translation work. In traditional English translation teaching, teachers often explain and analyze the knowledge points learned in the classroom. However, some translation knowledge points are abstract and complicated and difficult for students to understand, which leads to lower learning efficiency. Therefore, in order to better improve the teaching effect and cultivate high-quality English translators, it is necessary to use virtual technology and data mining methods to assist the establishment of corresponding links between translation methods and methods, and to achieve effective integration, thereby helping English translation Professionals perform language output work more quickly in the classroom. This article uses questionnaire survey and data analysis methods to explore the current status of existing English translation teaching models, and then uses virtual technology and data mining methods to deal with existing problems, thereby further enhancing the effect of English translation teaching and improving teaching quality. According to the survey results, most of the interviewed teachers and students believe that virtual technology has played a certain auxiliary role in the development of teachers' teaching work and students' English translation training, and has certain application value to the teaching mode of English translation.

Keywords: English translation · Teaching mode · Virtual technology · Data mining

1 Introduction

English translation is a highly comprehensive subject, which involves two aspects of language and background culture. This means that when learning English, students must not only master basic vocabulary, sentence patterns and grammar, but also learn the relevant background culture of China and English-speaking countries. In the traditional teaching mode of English translation, teachers are the leading role and students learn passively. The disadvantage of this teaching method is that it cannot arouse students' interest in reading. At the same time, due to the differences in the traditional language

environment, voice information processing methods and expression forms, it is difficult for some students to understand complex words, sentence structure features and application skills. Therefore, using virtual technology and data mining methods to explore the English translation teaching model can make the teaching process more vivid and visualized, thereby further improving the teaching model and improving the quality of teaching.

At present, there are endless researches on English translation teaching, virtual reality and data mining, and the research results are also very rich. For example, Wang Xue pointed out that using data mining algorithms to verify the new teaching concepts and methods of tea culture translation courses, so as to lay a good foundation for future tea culture translators [1]. Huang Jing believes that, to a certain extent, the education model combined with the Internet is beneficial to effectively increase the effectiveness of English translation classroom teaching [2]. Xu Min proposed that virtual reality technology should also be accepted by teachers and used in the teaching process, which is of great significance for English learning [3]. Therefore, this article uses virtual technology and data mining methods to study the teaching mode of English translation, keeping up with the trend of the development of the times, and has important research value and practical significance.

This article mainly elaborates from these aspects. First, it briefly explains what virtual technology is and data mining. Then, the teaching mode of English translation is elaborated. In addition, the application of decision tree algorithm in English translation teaching mode is also introduced. Finally, regarding the application of virtual technology in the English translation teaching model, a questionnaire survey was launched, and the survey results and analysis conclusions were drawn.

2 Virtual Technology, Data Mining Methods and Teaching Mode of English Translation

2.1 Virtual Technology

Virtual technology, with the help of computer networks, creates things that may or may not exist in objective reality, and is used to simulate scenes or things that may or may not exist in real life. At present, virtual reality technology has been widely used in various fields. Virtual technology already possesses the new characteristics of modern technology. It creates a real-time three-dimensional image world that reflects the changes and interactions of physical objects.

It has strong interactivity and can make language translation work more convenient. With the development of high and new technology such as computer network, communication engineering, artificial intelligence, and its powerful functions, people have conducted in-depth explorations in more fields of virtual reality. An important feature of virtual reality technology is immersion and experience. It puts people in a virtual world and produces immersive feelings that can improve people's work efficiency and help people to accomplish things that they could not do before [4, 5].

2.2 Data Mining

Data mining is the use of effective methods to extract useful and potentially valuable information from a large amount of incomplete, unstructured and semi-structured data information. Through the analysis of the original variables and statistical related knowledge, we can discover its inherent laws. The core of data mining technology is to unearth potentially useful information, so in order to achieve this goal, it is necessary to clarify what kind of original variables are being studied. Data mining technology is an interdisciplinary problem involving multiple disciplines and fields. It requires researchers to classify and integrate complex information from multiple perspectives into practical value and effective decision-making methods. The implementation process of data mining is shown in Fig. 1.

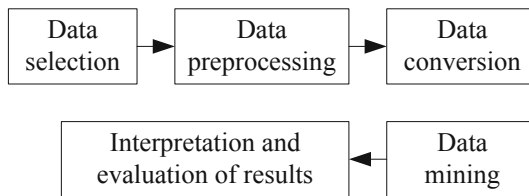


Fig. 1. The process of data mining

Data mining technology has its unique advantages in all aspects. First, it has strong applicability and high excavation efficiency. It can also reduce labor costs, improve work efficiency and productivity, so as to solve practical problems. The second is high flexibility and predictability. Through the use of these characteristics, the hidden relationships and rules can be found well, so that things can be predicted and analyzed, and then conclusions can be drawn. The third is strong operability. Ability to apply different types of knowledge points in different fields to deal with problems. But it requires professionals to be familiar with computer technology and related theories to get a solution [6, 7].

2.3 English Translation Teaching Mode and Research Status

The teaching model represents the integration of teaching philosophy and teaching practice. The formation of teaching mode also involves factors such as teacher composition, student combination, teaching conditions and evaluation. In the process of transformation, China's translation teaching model has mainly experienced these modes: the traditional translation teaching model, the translation teaching model after the transformation of the roles of teachers and students, the translation teaching model of multimedia computer-assisted teaching, and the integration of multiple disciplines and multiple perspectives translation teaching model.

The traditional English translation teaching model in China, whether it is a teaching model centered on translation skills, a teaching model centered on translation theory, or a teaching model combining the two, is that teachers dominate and students learn

passively, and the teacher introduces some translation theories and techniques, arranges relevant exercises, and students practice translation after class to achieve the purpose of consolidating theories and strengthening skills. This traditional teaching model, which emphasizes the text itself, ignores the interference of the learning subject and non-verbal factors, has a strong empiricism. Although this method allows teachers and students to communicate and interact, enhance the relationship between teachers and students, and improve classroom efficiency, it cannot solve the problem of language understanding well [8, 9].

At the same time, under the influence of structural linguistics, traditional English translation teaching focuses on the analysis of morphology and syntax. On the basis of comparing the grammatical structure of the language, teaching translation skills such as the order of English to Chinese. Translation education has thus changed from a traditional teaching method that focuses on smaller language units to a macro translation teaching method that emphasizes language, function, cognition and culture. For students, the teaching method that pays attention to the language microstructure of words, phrases, sentences, etc. cannot be ignored.

At present, China's translation teaching models mainly include these types of models. First is the mission mode. In this mode, the teacher does not directly pass the learning content to the students, but assigns tasks so that the students will gradually feel the learning content when performing the tasks assigned by the teacher. The student completes the homework, and the teacher provides feedback and feedback when the student completes the homework. Second is the situational teaching model. The situational teaching mode requires that it be separated from the traditional translation teaching and carried out in a real situation, and the simulated environment covers all aspects of social life. Third is the teaching mode under the multimedia and network environment. Taking advantage of the large storage capacity of the multimedia network, a translation database is created so that teachers can access various resources at any time. Multimedia technology can also record the results of each student stage so that teachers and students can test the effectiveness of their teaching and learning [10, 11].

2.4 Application of Decision Tree Algorithm in English Translation Teaching Mode

In the field of data mining, decision tree algorithm is one of the commonly used tools to classify and predict the researched matter. Decision tree algorithm has the advantages of high classification accuracy and simple generation model, which solves many problems in practical applications. Data mining, using its advantages in mining massive data, crawls a large number of English resources, including various network information resources and multimedia materials, and then obtains valuable information. In addition, by analyzing frequently used English translation vocabulary, sentence segments, etc., students can improve their understanding of teaching content and increase their interest in learning; and through corresponding technical means, students can strengthen their training in English translation [12, 13].

When the decision tree algorithm selects branch attributes on the root node and each internal node, the information gain is used as a measure. First, calculate the expected information of the data set, the calculation formula is as follows.

$$U(m_1, m_2, \dots, m_n) = - \sum_{g=1}^n O(d_g) \log_2(O(d_g)) \tag{1}$$

Then, the information entropy is calculated, which is later used to calculate the information gain.

$$F(B_e) = \sum_{r=1}^w \frac{m_{1r} + \dots + m_{nr}}{total} U(m_{1r}, \dots, m_{nr}) \tag{2}$$

$$U(m_{1r}, \dots, m_{nr}) = - \sum_{g=1}^n O_{gr} \log_2(O_{gr}) \tag{3}$$

Finally, calculate the information gain when B_e divides the data set.

$$Gain(B_e) = U(m_1, m_2, \dots, m_n) - F(B_e) \tag{4}$$

Among them, M is the node, d is the category, w is the number of different values, X is the data set, m_{gr} is the number of samples belonging to the category d_g in X_r , and O_{gr} is the proportion of the data samples of the category d_g in the subset X_r Proportion.

3 Questionnaire Survey on the Application of Virtual Technology to English Translation Teaching Mode

3.1 Questionnaire Design Process

The questionnaire survey selected 90 teachers and students from C University of Foreign Languages. Through the issuance of online questionnaires or paper questionnaires, the collection and quantitative analysis of the information filled in by users are carried out to draw conclusions of the questionnaire.

(1) In the preliminary preparation of the questionnaire, the number of questions should be as concise as possible to avoid fatigue of the interviewees.

(2) Questionnaire distribution. The questionnaire is distributed through online questionnaires, on-site questionnaires, and inviting friends to help ask friends and students around them to fill in the questionnaires. A total of 90 questionnaires are distributed, 90 valid questionnaires are returned, and the questionnaire recovery rate is 100%.

(3) Questionnaire analysis. The collected questionnaire information will be sorted out and the required information data will be obtained. Analyze the results of the questionnaire. The analysis results include the opinions of teachers and students on the application of virtual technology in English translation teaching mode. Some of the results obtained from the questionnaire are as follows.

3.2 Questionnaire Survey Content

The first part is the selection of 90 teachers and students from C University of Foreign Languages. The interviewees are divided into English translation teachers, English translation students and non-English translation students according to their occupations and majors, and their application of virtual technology to English translation. Investigate opinions in the teaching model.

The second part is to sort out the information collected in the questionnaire and understand the opinions of 90 citizens on the application of virtual technology to English translation teaching mode. Part of the questionnaire survey results are as follows.

4 Questionnaire Survey Analysis on the Application of Virtual Technology to English Translation Teaching Mode

In this questionnaire survey, 90 teachers and students from C University of Foreign Languages were selected, and the respondents were divided into English translation teachers, English translation students, and non-English translation students according to their occupations and majors. A questionnaire survey of opinions applied in the teaching model of English translation. The survey results are shown in Table 1.

Table 1. Attitudes toward application of virtual technology in English translation teaching mode

Project	English teacher	Students of English translation major	Students of non-English translation major
Very helpful	11	12	9
Helpful	9	8	11
General	8	7	9
Not helpful	2	3	1

As can be seen from Fig. 2, regarding the application of virtual technology to the teaching model of English translation, among the 90 interviewees, 32 people said it was very helpful, 28 people said it was helpful, and only 6 people said it was not helpful. It can be seen that most of the interviewed teachers and students believe that virtual technology has played a certain auxiliary role in the development of teachers' teaching work and students' English translation training, and has certain application value to the teaching mode of English translation.

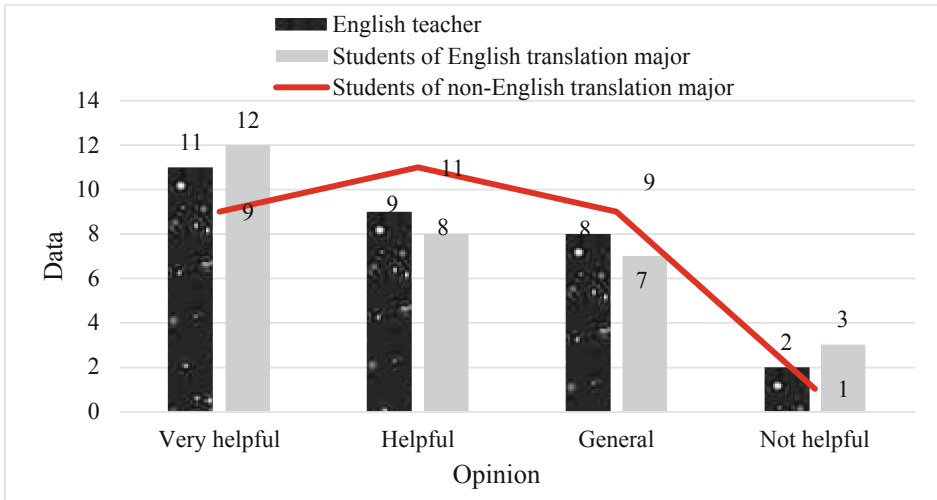


Fig. 2. Attitudes toward application of virtual technology in English translation teaching mode

5 Conclusion

From the current point of view, there are many problems in the existing teaching methods of English translation, which are mainly manifested in: teachers cannot grasp the content of textbooks in a timely and effective manner, lack the ability of independent innovation; learning time is short, which makes it difficult to keep up with the development requirements of the times, etc.. It is difficult for students to learn because of the single language types and simple expression methods used in the traditional teaching mode. With the help of virtual technology and data mining methods to solve these problems, it is feasible and operable, which can not only make English translation easier to understand, but also improve teaching efficiency to a certain extent. Therefore, this article uses virtual technology and data mining methods to conduct a preliminary discussion and research on the teaching mode of English translation, which has certain research significance and value.

References

1. Xue, W., Wei, Z.: Research on the TPACK mixed teaching model of English translation course of tea culture under wisdom education. *Fujian Tea* **041**(003), 259–260 (2019)
2. Jing, H.: Research on the reform and implementation methods of college English translation teaching model in the network environment. *Campus English* **474**(34), 26 (2019)
3. Min, X.: Research on the practical application of virtual reality technology in the constructivist English teaching model. *Mod. Vocational Educ.* **000**(020), 228 (2018)
4. Vieira Monteiro, A.M., Ribeiro, P.: Virtual reality in English vocabulary teaching: an exploratory study on affect in the use of technology. *Trabalhos em Linguística Aplicada* **59**(2), 1310–1338 (2020)

5. Pletz, C.: Which factors promote and inhibit the technology acceptance of immersive virtual reality technology in teaching-learning contexts? Results of an expert survey. *Int. J. Emerg. Technol. Learn. (iJET)* **16**(13), 248 (2021)
6. Kowalski, S., Samól, P., Hirsch, R.: Virtual reality tools in teaching the conservation and history of Polish architecture. *World Trans. Eng. Technol. Educ.* **18**(4), 399–404 (2020)
7. Lestari, M., Arono, A.: Evaluation of English teaching materials used at SD IT Ummi in Kota Bengkulu based on KTSP (Kurikulum Tingkat Satuan Pendidikan). *JOALL (J. Appl. Linguist. Lit.)* **3**(2), 88–102 (2019)
8. Siregar, M., Dewi, N.R., Lubis, S.: The development of learning material for seminar on English language teaching course based on flipped learning. *Budapest Int. Res. Critics Linguist. Educ. (BirLE) J.* **3**(1), 339–347 (2020)
9. Peji, A., Molcer, P.S.: Predictive machine learning approach for complex problem solving process data mining. *Acta Polytechnica Hungarica* **18**(1), 45–63 (2021)
10. Ruiz, R., Roubickova, A., Reiser, C., et al.: Data mining and machine learning for porosity, saturation, and shear velocity prediction: recent experience and results. *First Break* **39**(7), 71–76 (2021)
11. Papík, M., Papíková, L.: Application of selected data mining techniques in unintentional accounting error detection. *Equilibrium* **16**(1), 185–201 (2021)
12. Baloch, S., Samsani, S.S., Muhammad, M.S.: Fault protection in microgrid using wavelet multiresolution analysis and data mining. *IEEE Access* **9**, 86382–86391 (2021)
13. Buyrukolu, S.: Early detection of Alzheimer's disease using data mining: comparison of ensemble feature selection approaches. *Selcuk Univ. J. Eng. Sci. Technol.* **9**(1), 50–61 (2021)



Prediction Algorithm of Music Trend Based on Neural Network

Xuyan Wang^(✉)

Beihua University Teacher's College, Jilin 132013, Jilin, China

Abstract. Predicting music trends can allow users to understand current music trends, experience music and artistic atmosphere, and cultivate sentiment. The purpose of this article is to study a music trend prediction algorithm based on neural network. Explain the advantages of BP artificial neural network in predicting music trends. For the monthly, weekly, and daily golden music rankings of popular trends on the Ali music platform, the BP artificial neural network is used to predict the monthly, weekly, and daily playback volume. The selection of features is done through the artificial neural network BP model, the number of hearings and the average error rate of prediction. The average relative forecast error is 5.1%. Next, we compared the results of three methods to predict music trends: BP neural network model, quadratic exponential smoothing and ARIMA model. The experimental results show that the BP neural network algorithm can predict the development trend of data more accurately than other algorithms.

Keywords: BP neural network · Music popular · Popular trend · Prediction algorithm

1 Introduction

Music has undergone a huge transformation from traditional vinyl records to cassette music storage formats to CD storage and digital MP3 music storage, and music files have also evolved from self-storage to network storage [1, 2]. With the development of the Internet, the old music forms have gradually disappeared, and online music forms have become more and more common, and various music data related to them have also been created, such as downloading, listening, and commenting [3]. At the same time, Internet music users put forward higher IQ and personalized requirements for their music platform experience. How to better use various online music data services for music users is also a research hotspot [4, 5].

In recent years, due to the upsurge of neural networks and deep learning, many scholars have begun to apply it to prediction [6]. Some scholars have proposed a stock trend prediction method based on the convolutional neural network (GC-CNN) model of graph convolution features, which considers both stock market information and individual stock information [7]. Specifically, an improved graph convolutional network (IGCN) and a dual CNN are designed to construct GC-CNN, which can simultaneously capture stock market features and individual stock features. Six randomly selected Chinese stocks were used to prove the superior performance of the proposed GC-CNN

method. Experimental analysis shows that the method based on GC-CNN is superior to several stock trend prediction methods and stock trading strategies. Some scholars also showed a new method of Bitcoin trend prediction based on one-dimensional convolutional neural network (1D CNN) model [8]. First, they proposed a method to construct a useful data set that takes into account social media data, a complete blockchain transaction history, and many financial indicators. In addition, they proposed a cloud-based system characterized by an efficient distributed architecture, which allows us to use the above methods to collect large amounts of data to build thousands of different data sets. Using historical Internet music data, it is of great significance to study music trend prediction based on BP neural network [9].

This article introduces the source and research significance of music popular prediction, explains the main work and contribution of this article, and combines the organization structure of the thesis. The experimental data set comes from Alibaba Cloud Music Platform. Combined with the BP neural network algorithm, a new music trend prediction model is established, which can obtain accurate prediction results. According to the user's historical reproduction data, by predicting the number of songs that the artist will play in the target time, they recreate the artist who will become the trend, so that they can achieve precise control over the trend within a period of time.

2 Research on Prediction Algorithm of Music Trend Based on Neural Network

2.1 BP Neural Network

Neural network is a mathematical algorithm model that simulates the behavior of animal neural networks to process score information. The BP neural network is a multi-layer power supply network trained by the inverse error multiplication algorithm, and the optimization of the network weight factors has obvious characteristics [10, 11]. If the input and output are known, and the mechanism of extracting output from the input is not clear, the unknown process between entry and exit can be regarded as a "network"; and the entrants and exits of the network can be selected from the sample set to correspond to The sample is used as a "train" on the Internet. The input samples should be processed from the input layer, hidden layer and output layer [12]. When the exit is inconsistent with the ideal exit, the network enters the feedback process and uses the error between the two to adjust the weight between nodes to the ideal error level. After the training, when the network operator is provided again, the network will be produced according to the weight after the training.

The BP neural network algorithm is simple and the source is clear. It can identify samples of unstable factors or lack of factors, and match a large amount of input information. After learning, the indirect characteristics and rules of the sample can be assigned to the weight of the neural network connection. Even if some connections are wrong, the efficiency of the network will be slightly reduced, and the accuracy of learning is also high. After training, the analysis rules automatically derived from the network have strong nonlinear generalization of mapping and generalization, which can be used to predict unknown things. However, in practical applications, BP networks also have some

shortcomings, such as slow convergence, the choice of network structure and learning parameters also directly affect the ability to ensure convergence at the global minimum. Impact on network model optimization. These parameters are usually determined based on experience or the size of the error. Therefore, the training of the network has a certain degree of randomness and redundancy.

2.2 BP Artificial Neural Network Music Trend Prediction

BP artificial neural network can achieve fuzzy approximation of the nonlinear relationship between input and output. The combination of the S-shaped logarithmic or tangent activation function of the artificial neural network and the linear function makes it possible to simulate any non-linear relationship, and continuously explore the causal relationship between the input and output of popular music historical data, which can simulate popular music well. It is in a dynamically changing nonlinear system. In view of the fact that neural network data cannot achieve human-computer interaction, under the existing network technical conditions, it can be compensated by sufficient sample data, and when the characteristics of the problem and the number are transformed, the existing theory can be combined with MATLAB's The function function is designed to minimize the loss of information.

Neural network can make diversified predictions through self-programming in MATLAB. On the one hand, the neural network can regulate the value range in the prediction process by adjusting the training function category and the number of hidden layers in the program design. On the other hand, the artificial neural network can be realized by MATLAB. The characteristics and numbers of the actual problem are transformed through the conversion of input and output models, and the prediction problems that need to be solved through self-programming are explored. The learning samples can be predicted in the program according to actual needs. As a result, the design is easy to achieve the comprehensiveness of popular music prediction. In addition, the establishment of popular color prediction system and the development of application software are the inevitable trend of popular music prediction work.

3 Investigation and Research of Music Trend Prediction Algorithm Based on Neural Network

3.1 Data Collection

This article uses the music data of the Ali music platform for half a year as the original data, and then further processes these data to form a new data set, and then predicts from the perspective of regression prediction and time series prediction respectively, so that the singer's song is in the next Predictive study of the total number of broadcasts in the time period.

The data mainly contains the manipulated information of the song and the information of the singer to which the song belongs. The attributes included in the table are the song number, the account number of the singer corresponding to the song, the release time of the song, the initial number of songs played, the language of the song, and the gender of the singer.

The timestamp corresponding to the play_time attribute provided in the data set in this article can be accurate to the hour. For example, the timestamp 15688549 will be converted into a normal date format of 2021-10-30, 8:0:0, and the date value of 2021-10-30 is It is also provided in the collection_time attribute.

3.2 Evaluation Index

The error sum of squares used in this experiment is mainly used to measure the degree of fit of the quadratic exponential smoothing method, autoregressive moving average model, and BP neural network model for nonlinear fitting. The squared error is shown in Eq. 1:

$$E = \sum_{i=1}^n (\widehat{y_i} - y_i)^2 \tag{1}$$

In formula 1: E is the error sum of squares, Y_i is the i-th actual observation value, and n is the number of observation values.

This experiment needs to predict the daily audition volume of the artist’s songs, and the average relative error can be used to see the effect of the prediction simply and clearly. The average relative error is shown in Eq. 2:

$$E = \frac{1}{n} \sum_{i=1}^n \frac{|\widehat{y_i} - y_i|}{y_i} \tag{2}$$

In formula 2: E is the average relative error, Y is the i-th actual observation value, and n is the number of observation values.

4 Investigation and Analysis of Music Trend Prediction Algorithm Based on Neural Network

4.1 Building a Neural Network Model for Predicting Music Trends

When constructing the neural network model for the audition volume of the artist’s songs, the number of input neurons is 28, the number of output neurons is 1, the number of hidden

Table 1. Average error rate between predicted output and actual output

Time date	Average error rate
1–5	4.2%
6–10	5.7%
11–15	3.9%
16–20	6.6%
21–25	4.7%
26–31	5.5%

layer neurons is 6, and there is only one layer of hidden layer neurons. Constructing this algorithm model and predicting the artist’s daily audition volume of songs from October 1, 2021 to October 30, 2021 are shown in Table 1.

The above table records the output of the neural network algorithm from October 1, 2021 to October 30, 2021, the actual amount of audition of artist songs, and the average error rate of prediction. The average relative error of prediction is 5.1%. The bar graph that can be drawn based on the predicted data and the average error rate of the actual data in the above table is shown in Fig. 1:

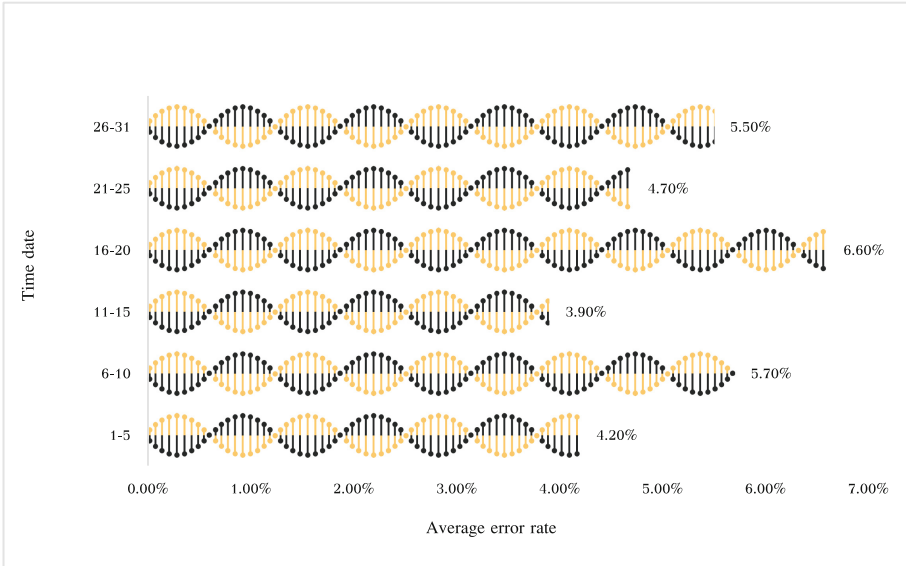


Fig. 1. Average error rate between predicted output and actual output

As can be seen from the above figure, the neural network can simulate the development trend of things to a certain extent when using the neural network algorithm to predict, and even some small data fluctuations can be predicted by the algorithm. The prediction error mainly depends on the training effect of the neural network model training data, the selection of factors influencing the amount of audition of artists’ songs, and the number of hidden neurons.

4.2 Comparison of 3 Prediction Algorithms

The preliminary screening method selects 5 artists, and uses the quadratic exponential smoothing method, the ARIMA model and the neural network algorithm to predict the amount of the artist’s song audition from October 1, 2021 to October 30, 2021. The predicted average relative error rate is shown in Fig. 2:

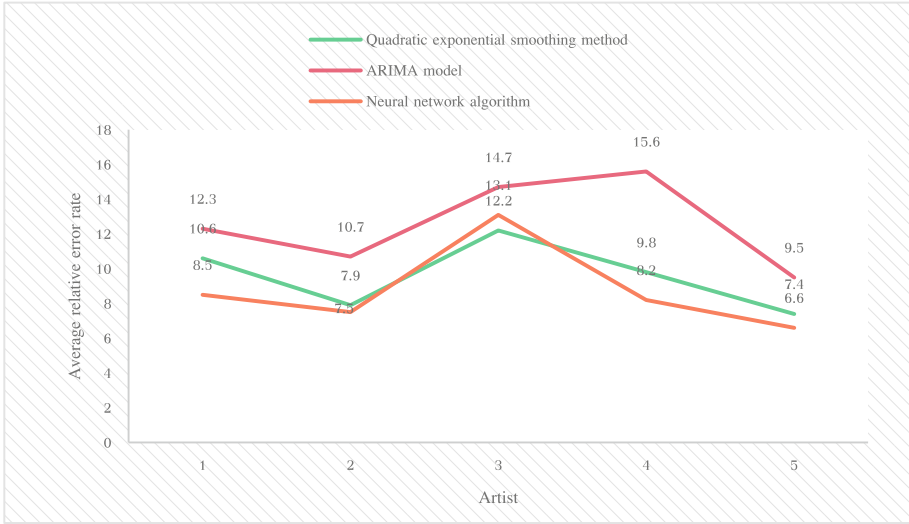


Fig. 2. Error comparison of experimental results

Since the prediction is the daily audition volume of artist songs from October 1, 2021 to October 30, 2021, there are 15 data in total, and the predicted time period is slightly longer, so the prediction effect is not very good. From the average value obtained in the above table, it can be seen that the prediction accuracy is the highest when using the BP neural network model.

Observing the line chart of the song audition volume of the artists from October 1, 2021 to October 30, 2021, and comparing them with the prediction results of these three algorithms, the following conclusions can be drawn:

Using these three prediction algorithms can predict the next stage of the artist’s song audition to a certain extent. But sometimes the prediction effect is that the quadratic exponential smoothing method is the best, sometimes the ARIMA model is the best, and sometimes the neural network algorithm model is the best, and no one algorithm can completely exceed the prediction results of other algorithms. But in general, the overall prediction level of the neural network algorithm is higher than the other two prediction algorithms; the use of the quadratic exponential smoothing method can find a linear function based on the past data to predict the future data. If the actual development trend of the predicted data is relatively stable and the fluctuation is small, the accuracy of the algorithm’s prediction is relatively high. On the contrary, if the fluctuation of the data is relatively large or the data has no long-term development trend, it is easy to cause relatively large prediction errors, and the prediction is The greater the distance, the greater the error; when using the quadratic exponential smoothing method for forecasting, the predicted value can only change along one trend. However, when using neural network algorithm to predict, it can reflect the various trends of the predicted value in a certain period of time. When the ARIMA model is used for prediction, the predicted value may show periodic trend changes due to the parameters of the model; compared with these

three prediction algorithms, the time used by the neural network model in the prediction time far exceeds the other two algorithms time. From the perspective of the development trend of prediction results, neural network algorithms can predict the development trend of data more accurately than other algorithms.

In practical applications, the BP neural network model, the quadratic exponential smoothing method and the ARIMA model can be used in combination to predict data.

5 Conclusions

In recent years, the application of neural networks in the field of prediction has gradually attracted attention, and there have been some gratifying explorations. This article is a research on the popular trend prediction of Ali music platform, mainly based on the characteristics of the music platform in the early stage to predict the amount of music played in the later stage. The content of this article can be divided into four parts: introduction, related algorithm description, feature selection model establishment, experiment and comparative analysis. The popular trend of music is mainly due to various uncertain factors before and during music playback. Because the internal relationship between the influencing factors is not clear, and the influencing factors are large in dimension, this article applies the latest popular BP artificial neural network to In the music trend forecasting, the music trend forecasting neural network is established, and the training model is adjusted to adjust the parameters to predict the music trend.

References

1. Hui, X., Sun, Y., Yin, F., et al.: Trend prediction of agricultural machinery power in china coastal areas based on grey relational analysis. *J. Coastal Res.* **103**(sp1), 299 (2020)
2. Maguluri, L., Rengaswamy, R.: An efficient stock market trend prediction using the real-time stock technical data and stock social media data. *Int. J. Intell. Eng. Syst.* **13**(4), 316–332 (2020)
3. Ghosh, I., Chaudhuri, T.D.: FEB-stacking and FEB-DNN models for stock trend prediction: a performance analysis for pre and post Covid-19 periods. *Decis. Mak. Appl. Manag. Eng.* **4**(1), 51–86 (2021)
4. Nipane, S., Kale, P., Kapsekar, P., et al.: Price trend prediction of stock market using outlier data mining algorithm. *Int. J. Innov. Eng. Sci.* **6**(4), 8–11 (2021)
5. Prachyachuwong, K., Vateekul, P.: Stock trend prediction using deep learning approach on technical indicator and industrial specific information. *Information* **12**(6), 250 (2021)
6. Lu, R., Lu, M.: Stock trend prediction algorithm based on deep recurrent neural network. *Wirel. Commun. Mob. Comput.* **2021**(7), 1–10 (2021)
7. Rubaie, E.: Crypto-currency price trend prediction using k-means clustering. *J. Eng. Appl. Sci.* **14**(9), 10594–10599 (2020)
8. Han, R., Song, D., Li, Y.: Sensor deviation anomaly trend prediction based on phase space reconstruction and support vector regression. *J. Phys. Conf. Ser.* **1930**(1), 012008 (2021)
9. Xue, Y., Feng, R., Cui, S., et al.: Traffic status evolution trend prediction based on congestion propagation effects under rainy weather. *J. Adv. Transp.* **2020**(7), 1–14 (2020)
10. Zhao, X., Li, W.: Trend prediction of event popularity from microblogs. *Future Internet* **13**(9), 220 (2021)

11. Thakkar, A., Chaudhari, K.: CREST: cross-reference to exchange-based stock trend prediction using long short-term memory. *Procedia Comput. Sci.* **167**(2020), 616–625 (2020)
12. Ni, H., Wang, S., Cheng, P.: A hybrid approach for stock trend prediction based on tweets embedding and historical prices. *World Wide Web* **24**(3), 849–868 (2021). <https://doi.org/10.1007/s11280-021-00880-9>



The Application of Panoramic Image Mosaic Technology in Modern Digital Painting

Tao Zhang^(✉)

Dalian Neusoft University of Information, Dalian, Liaoning, China

Abstract. With the development of the times, the increasingly mature digital painting technology has brought us into a new era of computer graphics. Digital painting, which inherited the spirit of traditional painting art, quickly entered the commercial market with its high efficiency, high expressiveness, and high tolerance, and supported each other with other digital visual arts, opening up a broader space for development. The geometric position transformation of a given digital painting can produce some interesting visual effects, which can profoundly affect the overall form and content of the digital painting. This article aims to study the application of panoramic image splicing technology in modern digital painting. On the basis of analyzing the classification of digital painting, the characteristics of digital painting and the basic theory of image splicing, in order to further confirm the application of panoramic image splicing technology in modern digital painting. The feasibility of this article is to verify the feature matching algorithm in panoramic image stitching technology. The experimental results show that the overall performance of the Delaunay grid feature matching algorithm is the best, with small fluctuations and low mismatch rate.

Keywords: Panoramic image stitching technology · Modern painting · Digital painting · Feature matching algorithm

1 Introduction

Art and scientific skills are two different forms that enable people to know themselves and reflect the entire objective world. The power of social progress based on science and technology can be achieved through extensive use and integration with other human societies [1, 2]. The mutual penetration of science and art has promoted the development of my country's culture and art and the revolutionary innovation of its forms. Technology has become an important carrier of contemporary social and cultural exchanges and contemporary artists' global intellectual and cultural exchanges, thus bringing a new way of thinking to contemporary artists' creation [3, 4].

The extensive application of panoramic image splicing technology in the development of digital painting art has differentiated traditional painting forms and digital paintings from each other, and has produced and developed an artistic expression form of panoramic image splicing-the art of panoramic image splicing [5, 6]. The extensive use of digital technology in the practice of modern painting art is not only reflected in

the materials and tools of painting, but also has an impact on the creativity and way of thinking of artists, causing changes in the creative process and expression techniques of painting, has produced a new type of artwork that is different from our traditional painting art form [7, 8].

Based on the analysis of the classification of digital paintings, the characteristics of digital paintings, and the basic theories of image splicing, in order to further confirm the feasibility of panoramic image splicing technology applied to modern digital paintings, this article analyzes the feature matching algorithm in panoramic image splicing technology.

2 Application of Panoramic Image Stitching Technology in Modern Digital Painting

2.1 Classification of Digital Paintings

(1) Classification according to the principle of image generation of digital painting

1) Bitmap

First of all, we can see through the idea that the basic unit of a bitmap is a pixel. Pixels, also called pixels, are the basic part of the image we see on the screen. When we look at the bitmap image under multiple magnifications, we can see that it is actually composed of many small color blocks, each of these small color blocks is a pixel, and the computer processes these points [9, 10]. Because of this calculation method and the idea of displaying bitmaps, its advantage is that it is easy to modify and create. It is a good idea to modify some color patches immediately. In addition, it can also create very sensitive images. If the artist constructs the canvas large enough, the color change can be accurate to every pixel.

2) Vector diagram

Vector graphics refer to images represented by geometric shapes based on mathematical equations such as points, lines, and polygons. In addition to bitmaps, another graphics mode. Although the bitmap creation function is very powerful, it also has a big disadvantage, that is, it cannot be zoomed in or out, because it consists of a certain number of pixels, so whether it is zooming in or zooming out, it will cause pixel loss. When you enlarge an image, the number of original pixels in the image is obviously not enough, so the computer will automatically copy the original pixels to enlarge it, but this mechanical copy and enlargement of pixels in equal proportions will cause the image to be obvious. It is not clear, there are jagged small squares [11, 12]. The vector diagram counteracts this problem well. Vector graphics do not record and calculate the color value of each pixel to form an image, but divide the image into geometric shapes one by one, and divide these geometric shapes into mathematics. Record and calculate the equation in the table so that no matter how many times the image is enlarged or reduced, the image is always clear, because the equation is constant, but the value is modified. Therefore, this creative method is very

common when designing large-scale advertisements and making huge posters. However, due to image segmentation and complex calculation methods, it cannot be used for very detailed creation. It is generally based on large pieces of paint, graphics, etc., which is very inconvenient for detailed painting. In addition, its creative process is far from traditional painting creation. Therefore, when creating digital paintings, artists will use a combination of bitmaps and vector graphics.

(2) Classified according to the composition of the painting space

1) Two-dimensional plane creation method

The so-called two-dimensional graphic creation method means that the creative concept is creation from level to level, that is, the creator starts with the method of creating a plane, and the final creation is also a plane artwork. It is characterized by easy creation, and the process is much simpler than 3D creation. Most digital artists prefer this method of creation. Because this way of creation makes them feel more free and relaxed, they can understand the inspiration of creation more quickly. But its disadvantage is that it is very difficult to create 3D multi-angles, and the spatial accuracy is not high.

(3) Three-dimensional creation method

The 3D creation method refers to the use of some 3D creation software to create a 3D model on the computer, and then through the design of the model, adding virtual lights, adding the scene, and outputting a virtual photo in the following form. In a 2D image or video. His idea is from three-dimensional to two-dimensional. Its advantage is that when creating a model, you can use the existing model to capture and design multiple angles, multiple scenes, and different actions. When creating large-scale, multi-angle and digital paintings, movies and TV, you can save a lot of work and make the effect more realistic. But there are also disadvantages, that is, the initial modeling process is very time-consuming and consumes energy. This method of creation is difficult to satisfy artists who need to quickly understand inspiration.

2.2 Features of Digital Painting

(1) The virtuality and non-uniqueness of the way of existence

In some cases, some accidental factors in the project are difficult to control. However, digital painting has been subverted at this point, and the new way of creation and dissemination is a breakthrough in art. It is created on the computer, the way to save the work is virtual and can be copied infinitely. Therefore, every piece of work we see is real, and every piece is a manuscript created by the author one by one. It is no different from traditional paintings in terms of beauty, composition, lines, colors, light and shadow.

(2) The creation process is convenient and quick

First of all, from the preparatory stage of painting, artists only need to use mouse, hand-painting board, pressure-sensitive pen and other equipment to create their

own creations on their computers, and they can truly achieve paperless painting. In addition, it is worth noting that many painting software currently have functions of erasing and modifying. This kind of button function is generally called the “regret button” by some professional artists. As the name suggests, the “regret key” means that when you make a mistake in a stroke, you can use the “regret key” to undo and redo, so that the artist can explore art with confidence, without worries, and break the irreversibility of traditional painting creation in the past, which greatly reduces the cost of artistic creation. These functions are also another progressive embodiment of digital painting.

2.3 The Basic Theory of Image Mosaic

(1) Image grayscale

Image gray-scale processing is the process of gray-scale processing an image. The images we shoot generally have color information, because in the field of digital image processing, most images only need grayscale images, and do not need images that contain color information. Because grayscale images also have rich information, including image edge information, image outline information and image texture feature information, and grayscale images occupies much smaller storage space than color images, which is conducive to image processing. The pixel value range of the grayscaled image is between 0 and 255, which is simple and does not involve color information.

(2) Image sharpening

Photos acquired through mobile phones and other photography equipment may have disadvantages such as low resolution, unclear images, and blurred images. At this time, image sharpening technology is needed to sharpen the edges of objects in the image to change the blurry place. To be clear, and at the same time make the image outline more clear. Image sharpening is mainly to enhance the contour features of objects, including linear sharpening and nonlinear sharpening. Image sharpening has been widely used in real life, including the sharpening of images by mobile phone software, the sharpening of human appearance by the monitor, and the sharpening of lesions by medical treatment.

(3) Image correction

When taking a picture, because the shooting device is not fixed or the shooting device is shaken, the captured photo will appear tilted. The process of correcting the tilted photo into a horizontal photo is called tilt correction. In order to merge and merge the text image, this paper chooses the method based on Hough line transformation to correct the skew of the text image, which makes the subsequent text image mosaic and fusion more convenient.

3 Experiment

The extraction of feature points in digital painting registration needs to be able to extract stable and high-precision feature points for the same object in different paintings when the imaging conditions change. At present, the positioning accuracy of feature points by

feature operators has reached the sub-pixel level. Therefore, the evaluation indicators for feature point extraction mainly include the following three points: 1) the repetition rate of point features extracted from overlapping areas; 2) extracted from paintings The number of feature points; 3) The time consumed for feature point extraction.

The repetition rate characterizes the proportion of the feature points that can be repeatedly extracted to all the extracted feature points, which is of great significance to the application of geometric parameter recovery using the feature points in the overlapping area. Repeatable feature points need to satisfy that the distance difference of feature points after homography-based transformation is less than a certain threshold. Set the feature point set of the reference image in the overlapping area as $\{A\}$, and the feature point set of the image to be registered in the overlapping area is $\{B\}$, the homography transformation matrix between the reference image and the image to be registered is H , then the repetition rate of the feature extraction operator is defined as:

$$R = \frac{Size\{(H * A) \cap B\}}{Size\{A\}} \quad (1)$$

Where $Size\{\}$ represents the number of elements in the set, and $Size\{(H * A) \cap B\}$ represents the number of feature points in set A that have corresponding feature points in B after H transformation. The repetition rate is an important indicator of the feature point operator. The feature point extraction operator with high repetition rate has better anti-interference performance against the environment, and the number of feature points extracted is larger, and it can also reduce the matching error during subsequent feature point matching.

In order to verify the feasibility of the algorithm in this paper, this paper will use the classic SURF algorithm and the improved ORB algorithm and the BRAND algorithm, and the matching algorithm based on the Delauay network and the filtering algorithm based on scale clustering proposed in this paper for comparison and verification.

4 Discussion

After the performance test of the image feature point extraction operator. The three feature extraction algorithms all have the ability to calculate the main direction of feature points, and have good stability against interference such as morphological transformation and noise.

In Fig. 1, the abscissa is the interference added to the reference image, the maximum value is normalized to 100, and the ordinate is the repetition rate of the point feature extraction operator, which is the result of averaging the repetition rates of the five types of interference. It can be seen from the figure that the BRAND extraction operator has the best performance, while the ORB algorithm is in the middle, and the SURF extraction operator has poor performance, mainly because after the feature transformation, the number of feature points is greatly reduced and the repetition rate is reduced.

In order to verify the effectiveness of the image matching algorithm proposed in this paper and the removal of false matching, for image feature extraction and matching, the experimental results and information are counted as a table. According to the definition, the recall rate of feature matching is calculated for comparative analysis.

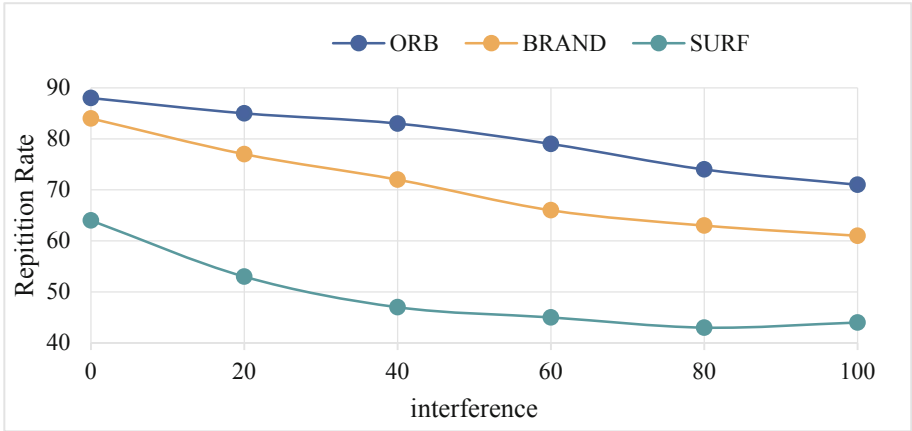


Fig. 1. Feature point extraction algorithm repetition rate curve

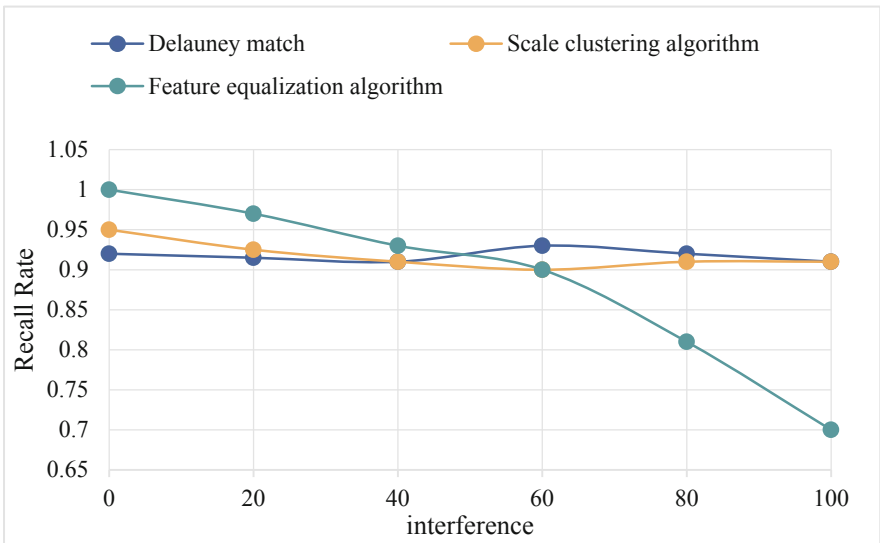


Fig. 2. Feature matching algorithm recall curvature

It can be found from Fig. 2 that the recall rate of the feature homogenization algorithm decreases faster under the condition of large interference, and the grid matching performance based on Delauney is the best, and the performance of the algorithm based on scale clustering is reduced compared to that.

It can be seen from Table 1 that the feature point offset and mismatch filtering of the traditional algorithm fluctuate greatly, and the matching algorithm based on scale clustering has a small number of matching pairs, but the mismatch filtering effect is obvious, the matching efficiency is high, and the feature Both the offset mean value and the offset variance value are relatively low, and the matching performance is better.

Table 1. Image feature matching result

Algorithm	Mean value of feature point offset	Variance of feature point offset	Matching logarithm	Logarithm of mismatch
Feature point distribution uniformization algorithm	107.626	2322.481	1158	61
Matching algorithm based on scale clustering	9.443	5.744	98	2
Feature matching algorithm based on Delauney grid	19.422	59.688	360	6

Based on the Delaunay grid feature matching algorithm, the overall performance is the best, with small fluctuations and low mismatch rate.

5 Conclusions

The digital age has truly arrived, and the digital society has become the objective environment for human survival today. The digital age has truly arrived, and the digital society has become the objective environment for human survival today. Every time a new art-related technology appeared in history, it inspired the original art to continue to exist and develop with a new look. Today, panoramic image stitching technology also introduces painting into fields that have not been explored before. All these new technologies can be called ways to expand the language of painting and provide new possibilities for the diversification and individual development of art.

References

1. Chilukuri, P.K., Padala, P., Padala, P., Desanamukula, V.S., Pvgd, P.R.: L, r-stitch unit: encoder-decoder-CNN based image-mosaicing mechanism for stitching non-homogeneous image sequences. *IEEE Access* **9**, 16761–16782 (2021)
2. de Lima, R., Cabrera-Ponce, A.A., Martinez-Carranza, J.: Parallel hashing-based matching for real-time aerial image mosaicing. *J. Real-Time Image Proc.* **18**(1), 143–156 (2020). <https://doi.org/10.1007/s11554-020-00959-y>
3. Abdukholikov, M., Whangbo, T.: Fast image stitching method for handling dynamic object problems in panoramic images. *Ksii Trans. Internet Inf. Syst.* **11**(11), 5419–5435 (2017)
4. Takeishi, N., Yairi, T.: Visual monocular localization, mapping, and motion estimation of a rotating small celestial body (Special Issue on AI, Robotics, and Automation in Space). *J. Robot. Mechatron.* **29**(5), 856–863 (2017)
5. Nocerino, E., Menna, F., Chemisky, B., Drap, P.: 3D sequential image mosaicing for underwater navigation and mapping. *CoRR abs/2110.01382* (2021)

6. Kim, D.T., Cheng, C.-H., Liu, D.-G., Liu, K.C.J., Huang, W.S.W.: Designing a new endoscope for panoramic-view with focus-area 3D-vision in minimally invasive surgery. *J. Med. Biolog. Eng.* **40**(2), 204–219 (2019). <https://doi.org/10.1007/s40846-019-00503-9>
7. Lati, A., Belhocine, M., Achour, N.: Robust aerial image mosaicing algorithm based on fuzzy outliers rejection. *Evol. Syst.* **11**(4), 717–729 (2019). <https://doi.org/10.1007/s12530-019-09279-4>
8. Sugiarto, E., Kurniawati, D.W., Febriani, M.: Computer-based art in folklore illustration: development of mixed media digital painting in education context. *IOP Conf. Ser. Mater. Sci. Eng.* **1098**(3):032017 (2021)
9. Pandey, A., Pati, U.C.: Image mosaicing: a deeper insight. *Image Vis. Comput.* **89** 236–257 (2019)
10. Maciura, L., Bazan, J.G.: Correction to: Granular computing in mosaicing of images from capsule endoscopy. *Nat. Comput.* **18**(3), 675 (2019)
11. Okaichi, N., Watanabe, H., Sasaki, H., et al.: Integral three-dimensional display with high image quality using multiple flat-panel displays. *Electron. Imag.* **2017**(5), 74–79 (2017)
12. Fernandez, M.A., et al.: Painting the Pacific: a comparative analysis of the lightfastness of watercolors made from indigenous plants in the Pacific region. *J. Health Disparit. Res. Pract.* **12**(4), 23–23 (2018)



Soccer Match Target Tracking System Based on Particle Swarm Optimization

Zerong Jiang¹ and Chen Zhu²(✉)

¹ Kunming University, Kunming, Yunnan, China

² Kunming Medical University, Kunming, Yunnan, China

Abstract. With the development of science and technology and the improvement of people's quality of life, the demand for video image processing technology is more and more extensive. Target detection and tracking in football game video is a very challenging task. The complexity of its moving background and picture makes football detection and tracking become the difficulty of image processing technology. This paper aims to study the football game target tracking system based on particle swarm optimization algorithm. Based on the analysis of the principle of image preprocessing and target tracking, combined with particle swarm optimization algorithm, a target tracking system is designed and implemented. In order to verify the effectiveness of particle swarm optimization algorithm, this paper takes a football game video on the Internet as a test sample to compare the tracking effects of particle swarm optimization algorithm and classical target tracking algorithm. The experimental results show that compared with the mean shift algorithm, the particle swarm optimization algorithm in this paper has more correct frames and better tracking effect.

Keywords: Particle swarm optimization · Football match · Target detection · Target tracking

1 Introduction

Video target tracking technology has always been one of the most hot research topics in the field of computer vision applications, and has attracted widespread attention at home and abroad [1, 2]. Video target tracking is a process of constantly guessing the target orientation in the video through computer technology. Its main task is to identify the target through each frame of the video and establish a trajectory for the target, so as to finally locate the target [3, 4].

Nowadays, many researchers are committed to developing target tracking algorithms, and the theory has made rapid progress. Although many innovations have been made, there are still many challenges to target tracking. The characteristics of the algorithm are affected by many reasons. At present, there is no algorithm that can meet the target tracking in different video occasions. Therefore, the problems in specific application fields often need to be solved according to the technical characteristics of specific application fields. Some scholars summarized the current research progress in the field of

target tracking, summarized the algorithms provided, introduced the characteristics and studied the online learning methods [5, 6]. At the same time, he also believes that a good target tracking system needs excellent robustness, adaptability and real-time. Some researchers directly use the RGB element histogram to obtain the whole field information. First, they separate the image to obtain the player's position, then use the pattern matching method and Kalman filter to track the players, and finally use the histogram to deal with the occlusion of the players, which has poor computational stability [7, 8]. Other researchers use DPM detector to determine the individuals in the map, divide the measured individuals into three groups, including team A, team B and others, finally get the general probability results, and establish the conditional random field model [9, 10]. On the basis of fusing the coordinate data of multiple cameras, some scholars use the pictures taken by multiple cameras installed on the game site, establish the observation mode through the strategy, and project it at the level of the game site. In this way, a multi-modal two-way probability function is obtained, which represents all possible situations of players on the court plane [11]. Some scholars also proposed a multi view method based on SIFT feature for multi person tracking, and assumed that the future ball game video is recorded by a high-resolution camera with multi view feature, so as to provide a free online viewing experience. Accurate 3D player information can also track the target more effectively, and use the 3D signals shared by all cameras to deal with the occlusion problem. The scholar has shown through practice that the tracking result is more accurate than the traditional sift method based on 2D signals. To sum up, target tracking in football game video has achieved some results.

On the basis of consulting a large number of domestic and foreign references related to target detection, combined with the principles of image preprocessing and target tracking, this paper designs a target tracking system. The system mainly includes four modules: target HSV feature extraction module, median filter module, target detection module and target tracking module based on particle swarm optimization algorithm. After realizing these four modules, this paper tests the tracking effect of the algorithm.

2 Research on Soccer Match Target Tracking System Based on Particle Swarm Optimization

2.1 Image Preprocessing

(1) Image denoising

Image noise mainly affects the random signal formed in the process of image data acquisition and image propagation, which is not conducive to human recognition of useful signals. Noise can also be regarded as a multidimensional random process, so a random process should be used to describe image noise. Image denoising is also a key content in image preprocessing. Image smoothing or filter is also an important technology to eliminate random noise in image.

(2) Morphological image processing

Morphology is the most common image processing task completed according to the graphic characteristics of image processing. Because the morphological operation uses the structured parameters of the input graph, it produces the output function

of the same size. In morphological mode, if the eigenvalues of all level variables in the output image are based on the contrast between the corresponding graphics in the input image and its adjacent images. Then by determining the length and type of neighborhood, a morphological method more sensitive to the special graphics in the input image can be established.

The most basic morphological task is expansion and corrosion. Expansion is to add pixels to the target boundary in the graph, while corrosion is to subtract pixels from the target boundary. Adding or removing image features from an image depends on the size and form of custom parameters used to process the image. Under the multimodal magnification and erosion model, the characteristics of all pixels in the output image are defined by the standards of corresponding and adjacent images in the input image. The standard of image processing is to define the expansion or corrosion in the operation.

2.2 Principle of Target Tracking

(1) Establishment of target model

Because the motion of the target can not be accurately estimated and predicted, the approximate estimation method is used to establish the target model. Target motion can be divided into two modes: constant speed and variable speed, as well as linear and nonlinear modes. Various possible situations need to be considered when building the target model, which is usually called global statistical model. However, this global statistical model meets all requirements, and in some cases, its adaptability or accuracy is not high. Therefore, the research on the dynamic characteristics of the current time is gradually carried out to predict the next time of the target. The model based on this idea is called “current” statistical model.

(2) Target detection method

1) Based on the color information method, the color information of the target information is measured to obtain the target color features. RGB, HSI, HSV and other color spaces are widely used in image processing. Among them, RGB is the main color space, which can be converted to other color spaces. In practical application, the color histogram model can be created by collecting the color information to be tracked, because color is the most intuitive, visible and reliable feature of image shape, color and texture. At the same time, target tracking often uses the target color histogram as the important information of detection, because the change of target size, direction and speed will not have a great impact on the target color.

2) The method based on pattern matching is to create a set of patterns through the prior knowledge of the tracked object, and calculate the correlation or difference between the fixed template and the candidate image region in the image space to identify the tracking target.

3) Based on the statistical learning method, different learning statistics or neural network methods are used to prompt the characteristics of the target, and different clusters are formed in the image space to receive the target. The target

recognition method based on neural network uses the learning and classification ability of neural network for feature extraction and target recognition. Input more correct samples and modify the parameters of the neural network algorithm so that the samples can be correctly identified when inputting specific samples.

(3) Target tracking method

1) Centroid tracking

When the target image is regarded as a thin sheet with uniform density distribution, the center of gravity of the thin sheet is the centroid of the target image. The centroid is usually a point in the center of the target graph. Even if the azimuth of the target changes, the displacement change at this point is quite small, which makes the centroid tracking method more stable. Because of its strong clutter suppression ability, this method is widely used in the field of target tracking.

2) Center of gravity tracking

The center of gravity, also known as the center of mass of an object, is the center of the static moment of matter on a certain axis. Assuming that the target image is regarded as a thin plate object with inconsistent mass density, and the gray value of each pixel in the target image is defined as the mass density of each point in the target image, the increase of centroid can be estimated by centroid definition method.

3) Edge tracking

There are several ways to select tip tracking points. It can take two edges, such as edge turning points or prominent corners, left and right edges, or top and bottom points. The edge monitoring method is simple and easy to apply. However, this method requires the contour of the target to be clearer and more stable. The holes and gaps in the image will make the stop of the tracking point unstable. Therefore, edge tracking will be greatly affected by noise.

4) Match tracking

Matching tracking requires a “standard” a priori target and will be detected in the tracking track. In the process of tracking, the target region must be determined first, and the target attribute extracted from the target region is the next attribute vector. For example, you can select the gray value or histogram of the target area. In each subsequent image frame, each region of the image is searched to find the position most similar to the target, that is, the target is considered.

3 Experiment

3.1 Extract Target HSV Feature Module

RGB model is the most widely used color model, which may be consistent with the strong perception of red, green and blue by human eyes, but it is not suitable for the intuitive interpretation of color. HSV model uses shadow and saturation, and the description of objects with values is more in line with the intuitive perception of human vision. Compared with RGB model, the correlation between each channel in HSV model is very small. Therefore, this algorithm obtains the target attribute vector by exporting HSV attributes.

3.2 Median Filter Module

Because the collected video image is inevitably accompanied by some noise interference, the noise reduces the image quality, blurs the image, and even drowns the feature information, which brings difficulties to the subsequent image processing and analysis. Therefore, median filtering aims to improve image quality and derive image features. In view of the large amount of data required by median filtering and the need of repeatedly using data, the same function can be applied to each pixel in the image. The algorithm is one of the most time-consuming links in the system and affects the running speed of the system. Therefore, this design adopts grouping comparison technology, avoids many comparison functions, and reduces the area and energy consumption of the design module.

Median filter replaces the value of a point with the median of all points near a point. This neighborhood is generally called window. In median filter image processing, we must first select the window w containing odd even point values, scan the window on the image, arrange the values contained in the window, and finally obtain the intermediate value as the node chromaticity value, so as to reduce the noise in the image. The median filter can be expressed by the following formula:

$$g(x, y) = \text{median}\{f(n - k, m - 1)\} \quad (1)$$

Where $g(x, y)$, $(m-k, n-1)$ is the chromaticity value of the pixel.

3.3 Target Detection Module

- (1) The gray converted video image is first filtered to remove noise. On this basis, the original background model is established;
- (2) Next, the background difference is created by using the current video image and the updated background, and the background model is updated in real time according to the result of the background difference, with the original background difference as the initial reading;
- (3) The difference is performed again by using the background difference value and the threshold value, and the binary motion information is received according to the difference result;
- (4) Finally, the binary motion information is morphologically processed to further connect the motion region;
- (5) The moving target is segmented according to the motion deterministic algorithm.

3.4 Target Tracking Module

When tracking a moving target, it is very sensitive to the factors of the target itself and the environment. The size and shape of the moving target, whether the target is rigid or non rigid, whether the target is deformed, whether the target is moving smoothly or violently, whether the target background is simple, whether the camera is stationary or moving, whether the tracking target is single or multiple tracking, etc. The gray features of the image are highly stable and have little effect on the distortion or rotation of the target. Therefore, when acquiring the target template, this paper uses the gray feature of the image.

Assume that the gray level of the image is m ($0-m-1$). Generally, M is equal to 256, a total of 256 Gy levels. The histogram function of the target template can be expressed as follows:

$$q = \{q_u, u = 0, 1, \dots, l - 1\} \quad (2)$$

$$q_u = \frac{1}{n} \sum_{i=1}^n \delta[d(x_i) - u], u = (0, 1, \dots, l - 1) \quad (3)$$

Where, $d(x_i)$ is the gray feature mapping function of points x_i . The specific target tracking process is shown in Fig. 1:

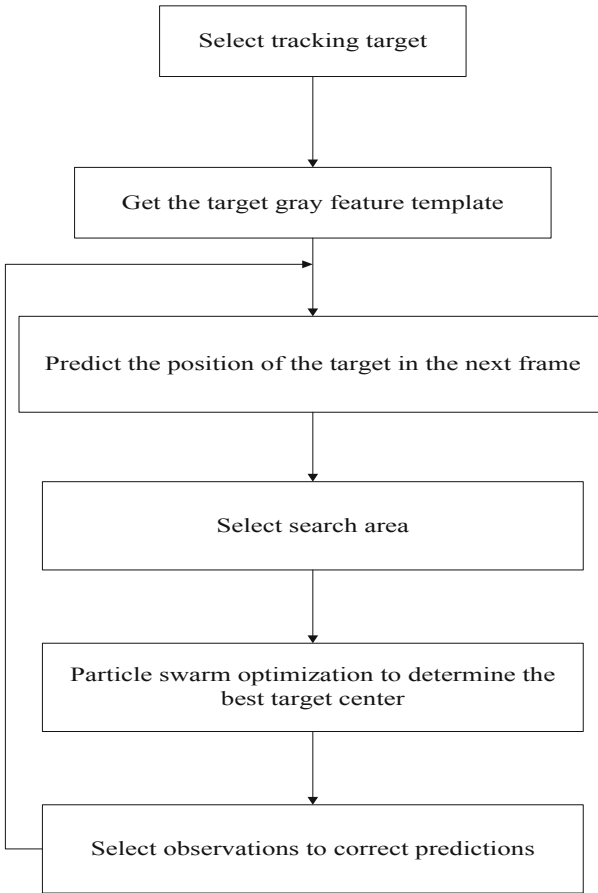


Fig. 1. Specific target tracking flow chart

4 Discussion

In order to verify the effectiveness of the algorithm, this paper takes a football game video on the Internet as a test sample. The video is 40 min long and a total of 23645 frames. The mean shift algorithm is selected to experiment with the experimental data, and the feasibility of the algorithm can be proved by comparing with the particle swarm optimization algorithm in this paper. The experimental results are shown in Table 1.

As can be seen from Fig. 2, the accuracy of mean shift algorithm for football tracking is 68.8%, and the accuracy of this algorithm is 90.4%. Therefore, compared with the mean shift algorithm, the particle swarm optimization algorithm in this paper has more correct frames and better tracking effect.

Table 1. Tracking results of different algorithms

Method	Total number of frames	Track the correct number of frames	Tracking accuracy
Mean-shift	23645	16275	68.8%
The algorithm of this paper	23645	21373	90.4%

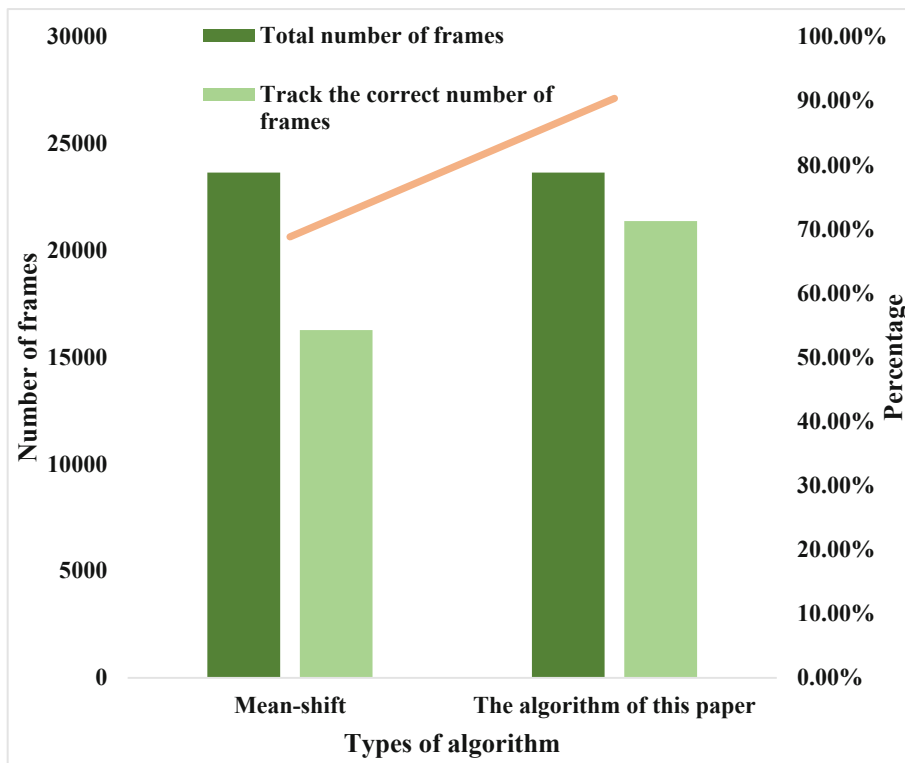


Fig. 2. Tracking results of different algorithms

5 Conclusions

As a kind of video with a large audience, sports video has also attracted more and more attention of researchers. Football game is the most concerned sports event, and there is more and more demand for football video analysis and processing. In this paper, the existing detection and tracking technologies are introduced and analyzed in detail, and a target tracking system is designed combined with particle swarm optimization algorithm. After experimental verification, this algorithm realizes football tracking well.

References

1. Bai, X., Wang, Z., Lei, Z., et al.: Target tracking for wireless localization systems with degraded measurements and quantization effects. *IEEE Trans. Industr. Electron.* **65**(99), 9687–9697 (2018)
2. Dong, P., Jing, Z., Leung, H., et al.: Student- t mixture labeled multi-Bernoulli filter for multi-target tracking with heavy-tailed noise. *Sig. Process.* **152**, 331–339 (2018)
3. Li, Z., Chen, X., Zhao, Z.: Design of standoff cooperative target-tracking guidance laws for autonomous unmanned aerial vehicles. *Math. Probl. Eng.* **2021**(2), 1–14 (2021)
4. Hou, L., Lian, F., Tan, S., et al.: Robust Generalized labeled multi-Bernoulli filter for multi-target tracking with unknown non-stationary heavy-tailed measurement noise. *IEEE Access* (99), 1 (2021)
5. Fang, X., Chen, L.: Noise-aware manoeuvring target tracking algorithm in wireless sensor networks by a novel adaptive cubature Kalman filter. *IET Radar Sonar Navigat.* **14**(11), 1795–1802 (2020)
6. Liu, B., Tharmarasa, R., Jassemi, R., et al.: Extended target tracking with multipath detections, terrain-constrained motion model and clutter. *IEEE Trans. Intell. Transp. Syst.* (99), 1–17 (2020)
7. Wang, Y., Wang, X., Cui, N.: Target tracking in boost stage with quantized measurements of space-based infrared cameras. *Trans. Japan Soc. Aeronaut. Space Sci.* **63**(5), 222–232 (2020)
8. Garcia-Fernandez, A.F., Ralph, J., Horridge, P., et al.: A Gaussian filtering method for multi-target tracking with nonlinear/non-Gaussian measurements. *IEEE Trans. Aerospace Electron. Syst.* (99), 1 (2021)
9. Li, S., Cheng, Y., Wang, H., et al.: Distributed multisensor multitarget tracking algorithm with time-offset registration. *Xibei Gongye Daxue Xuebao/J. Northwestern Polytech. Univ.* **38**(4), 797–805 (2020)
10. Mulimani, N., Makandar, A.: Sports video annotation and multi- target tracking using extended gaussian mixture model. *Int. J. Recent Technol. Eng.* **10**(1), 1–6 (2021)
11. Chen, T., Yang, P., Peng, H., et al.: Multi-target tracking algorithm based on PHD filter against multi-range-false-target jamming. *J. Syst. Eng. Electron.* **31**(5), 859–870 (2020)



Phase Study of CT Respiratory Slice Constructed by Three-Dimensional Image Fusion Technology

Hanxu Zhang^(✉)

The First Affiliated Hospital of Hebei North University, Zhangjiakou 075000, China

Abstract. Human absorption refers to the process of inhaling air or carbon dioxide from the inside to the outside of the body. It is an indispensable main component in human life activities. In nature, humans contact food through the lungs and inhale gas and air from the body. When people enter the respiratory tract (including skin, throat and other parts), they will cause dizziness to varying degrees; If not handled timely and effectively, it may lead to death or even disease. However, the blood released by breathing will contain a large number of free blood clots and hemoglobin, resulting in the occurrence of hypoxia and poisoning of human tissues and the increase of disability events. Therefore, based on three-dimensional image fusion technology, this paper constructs the mathematical model of CT respiratory slice. Firstly, this paper introduces the imaging principle of CT image and the characteristics of the image, then constructs a three-dimensional image fusion model and tests the error of the model. Finally, the test results show that compared with the traditional algorithm, the proportion of matching point pairs increases, and the registration error is less than that of the traditional algorithm. In terms of running time, it proves the feasibility of this method.

Keywords: 3D image · Image fusion technology · CT respiratory slice · Phase research

1 Introduction

CT technology uses computer to establish the three-dimensional structure of human voice, blood oxygen saturation and air pressure. It can detect the breathing area of human body in real time, and observe its position within the visual distance, so as to determine whether there is lack of air in the respiratory slice. CT technology is more and more widely used in the medical field, and its effect on disease diagnosis and treatment directly depends on the image fusion algorithm [1, 2].

Many scholars have done relevant research on 3D image fusion technology. Some scholars believe that respiratory slice refers to a characteristic section that can directly contact the human body and form a specific shape and structure under the action of a certain force. It is composed of multiple points and has the same section at different positions, and these characteristic sections are irregular geometric figures, which contain

a large number of complex and changeable and numerous, Therefore, it can be analyzed and processed by computer software [3, 4]. Other scholars used three-dimensional computed tomography to measure the wall thickness (WT) and airway area (AI) of the third to sixth generation bronchus, and the percentage of low attenuation area of lung less than - 950 Hu (% LAA). To evaluate the image quality of bone vascular fusion rendering (VR) images reconstructed by three-dimensional “black bone” magnetic resonance imaging (MRI) based on steady-state acquisition cycle phase (fiesta-c) sequence and time-of-flight magnetic resonance angiography (TOF-MRA) [5, 6]. The above research has laid the foundation for this paper.

This paper will study the two-dimensional representation of moving and stationary bodies by using machine learning system and establish the corresponding mathematical model. This method can quickly and accurately obtain the corner position of the patient’s proximal intravascular respiration, so that the doctor can observe the surrounding environment of tumor disease from multiple angles, and provide an important basis for surgical evaluation and decision-making.

2 Discussion on Constructing Phase of CT Respiratory Slice Based on Three-Dimensional Image Fusion Technology

2.1 CT Image

2.1.1 Imaging Principle

CT is the abbreviation of computed tomography, which also represents computed tomography. It is an image reconstruction technology. The work flow can be summarized as follows: first scan a structural longitudinal section of a three-dimensional object, and then obtain the data of the structural cross section of the object. After obtaining these data sets, then according to some mathematical principles, reverse calculation is carried out to obtain the main parameters consistent with the cross-section structure of the object, and finally restore the longitudinal section. By using display technology, the transverse section of the object can be obtained in the main parameters. The image is a sectional image on which various calculations are performed by the computer. This technology is applied to the medical field. Although similar mathematical principles are applied to image reconstruction, computed tomography is also divided into different types according to different sources and attribute (i.e. physical principles) parameters constituting image reconstruction. Due to different physical principles, the functions and application fields of these CT, such as CT, mrt-ct, r-ct, u-ct, etc. are also very different [7, 8].

2.1.2 CT Image Features

The imaging principle of CT image is based on the gray information represented by various human tissues (including normal and abnormal tissues) when absorbing X-rays, that is, pixels (the basic unit of CT image). CT images scanned by different CT machines have different sizes and pixel numbers [9, 10]. The smaller the pixel, the larger the number, and the more detailed the image, that is, the higher the spatial resolution. CT image is a sectional image obtained by computer scanning, but because different

human tissues have different attenuation coefficients relative to X-ray, the gray value of CT image is different. If the density of human tissue is high, the X-ray attenuation will also increase, that is, the gray value of CT image will increase, otherwise the gray value of CT image will decrease. According to medical regulations, if the CT gray value is set to 0, the gray value of CT image is obtained by using the conversion method similar to the temperature measurement unit according to this standard, and the unit is Hu. In other words, CT value is the relative density of human tissue, expressed in H, and its calculation formula is as follows:

$$H = 1000 \times \frac{\mu - \mu_{H_2O}}{\mu_{H_2O}} \quad (1)$$

2.2 CT Respiratory Slice Image Preprocessing

Pictures are also the main means of information visualization. Through a medical picture, it can bring auxiliary inspection and processing signals to medical personnel intuitively and vividly. In particular, experienced radiologists and clinicians can get a lot of useful information from this picture. However, due to the influence of many factors such as image equipment and acquisition conditions, the original image generated from the scanning device can also produce the decline of image quality and even artifact phenomenon. This leads to the inability of quantitative evaluation of image quality in medicine [11, 12]. Therefore, the main task of post-processing medical images with electronic computers is to improve the signal-to-noise ratio of captured medical images, that is, to improve the image quality. It eliminates the noise and influence in the image, highlights the range or boundary of the object of interest, and is applied to edge extraction and image segmentation, which lays a foundation for further division and reconstruction. The main purpose of preprocessing is to obtain volume data field with good three-dimensional visualization effect by using advanced two-dimensional image processing technology. Effective preprocessing and tissue separation is an important guarantee to accurately reveal the internal structure of human tissue and lesion information through three-dimensional image processing and fusion.

Compared with ordinary images, medical images have inherent fuzziness and non-uniformity. Medical image processing refers to the process of extracting, analyzing and studying the biological features containing targets by using computer technology, and transforming them into specific functional information. It includes image preprocessing (i.e. appropriately cutting the original data to make it a digital signal with certain significance and can be recognized. Therefore, preprocessing the collected medical image data is an extremely important task.

2.3 Image Processing Method

For all pixels in the image, take a certain area as the central point, and replace the gray value of the pixel with the weighted average of the gray value of each pixel in the area, which is the neighborhood average method. The specific method is to select a square area called window or smoothing mask, which is a two-dimensional weight array. The

smoothing step includes dragging the window on the graph, and then Face the image in the window and update its gray level through expression 2. After each image is scanned once, the smoothing of the image is turned off. This is also the easiest way to smooth the image in space. The general expression of the weighted average algorithm is:

$$g(x, y) = \sum_{i=-k}^k \sum_{j=-k}^k w(i, j)f(x + i)(y + i) \tag{2}$$

Where, (x, y) is the central element of the window; $F(x + I)(y + J)$ is a pixel of a noisy image; $W(I, J)$ is the weighted value, and its window size is $(2k + 1)(2k + 1)$.

2.4 3D Image Fusion Process

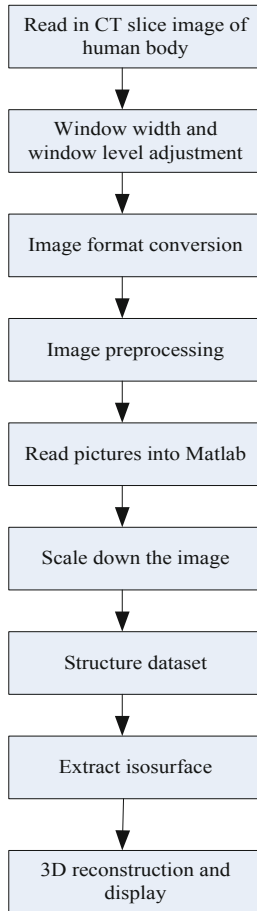


Fig. 1. CT 3 D reconstruction process of images of respiratory sections

Figure 1 is a three-dimensional image fusion process. The three-dimensional image fusion of medical images involves using a series of two-dimensional sectional images to reconstruct the three-dimensional image model. Firstly, all the two-dimensional sectional images required to form the three-dimensional model are read one by one in the PC. Then convert the image format from DICOM format to BMP format to prepare for VC processing. Process images in VC environment as needed. In order to speed up the execution of the program, the size of the image, that is, the size of the matrix, is reduced accordingly according to the hardware conditions of the computer. Then read the image sequence into Matlab to prepare the structural data, and then all representatives of the data set.

3 Experiment

3.1 CT Respiratory Slice Image Data Acquisition

The essence of 2D graphics visualization fusion technology is to decompose and process the 2D data formed by these devices to obtain the spatial graphics of 3D objects. The implementation is from two-dimensional to three-dimensional. First, we must deal with the overlapping relationship between two-dimensional data and three-dimensional spatial data. In fact, the number of volumes with three-dimensional spatial information is a collection of two-dimensional tomography data composed of several continuous sections. The results also show that in order to achieve better reconstruction effect, at least nearly 100 cutting data are needed for reconstruction. The more original the data, the better the 3D reconstruction effect. Finally, each pixel of these three-dimensional images is filled with the two-dimensional image data. The more complete the image data is, the more realistic the effect is, but the data processing takes time and the occupied space will increase.

When selecting 3D image fusion data, we should not only reconstruct the complete 3D image under given conditions, but also consider the software memory and the amount of calculation required in the reconstruction process. At most, as little image data as possible can be extracted. Based on this amount, a satisfactory three-dimensional effect can be reconstructed.

If the data from medical imaging equipment can be transmitted and managed on an ordinary computer, the computer shall transmit and digital medical images in accordance with DICOM specifications and convert them into corresponding data formats. The standard communication protocol of medical equipment is an industry standard for computer to computer communication of digital information and related medical images.

3.2 Experimental Process

The three-dimensional reconstruction of respiratory slice CT image used in this work is a fast three-dimensional modeling based on the MC algorithm of Mimics software. Mimics software can process and convert large format two-dimensional scanned images (such as CT and MRI images) on a personal computer, and create a three-dimensional digital model for editing and processing. Import CT images in DICOM format into

Mimics software. MICs software can directly read DICOM images. The converted CT image will automatically generate three different views: upper left, lower left and lower right. After importing the image, the direction is also set automatically. However, some files must be manually defined to determine the orientation of the image. The model can be positioned in six directions, namely up, down, left and right, front and back, which are called up, down, left, right, front and rear in medicine. Because these three different views have certain correlation, CT images can be located quickly.

4 Discussion

4.1 CT Respiratory Slice 3D Model Algorithm Error Analysis

Table 1 shows the error analysis data of this algorithm and traditional algorithm.

Table 1. Algorithm error data

	The algorithm of this paper	Traditional algorithm
Feature points	1856	2065
Matching points	1468	1658
Registration time/s	32.36	52.3
Registration error/cm	2.36	3.25
Storage format	PCD	PCD

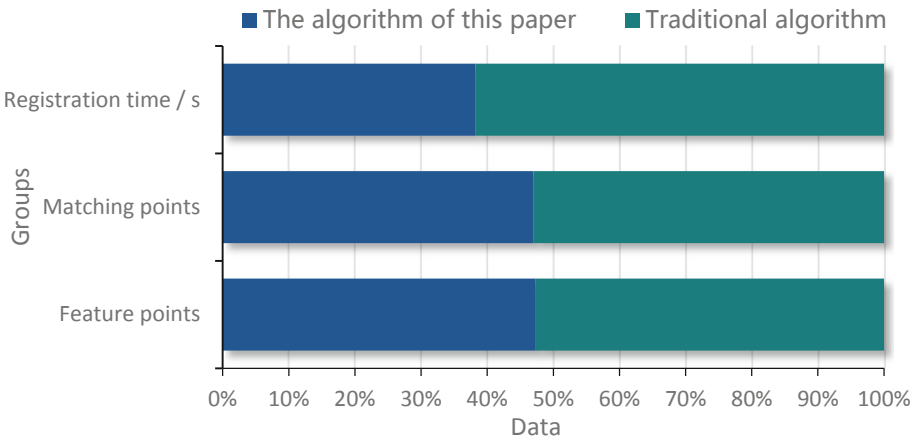


Fig. 2. Algorithm error comparison

Initializing from the template collection is critical to completing the shape. The collection of 3D models similar to the input object can provide a rough estimate of the overall shape of the approximate input image, and the elements of symmetry and

visibility jointly limit the completed results to similar topology and geometry to the input data, while reducing the input depth noise mode. Visibility constraints eliminate topology errors, symmetry energy improves overall high-level consistency, and reduces artifacts caused by visibility checking. As can be seen from Fig. 2, compared with the traditional algorithm, the proportion of matching point pairs increases, and the registration error is less than that of the traditional algorithm. From the perspective of running time, it proves the feasibility of the method in this paper.

5 Conclusion

In recent years, with the progress of science and technology, three-dimensional imaging technology has also been widely used in the field of medical and health care. Firstly, a CT respiratory slice model composed of two images in fast two-dimensional fingerprint is constructed by using CT and coherent synthesis principle. Then, according to the analysis results, a simple, effective and intuitive visual measurement is established, and the optimization results of this algorithm for 3D image fusion technology to construct CT respiratory slices are obtained.

References

1. Abugabah, A., Mehmood, A., Al Zubi, A.A., Sanzogni, L.: Smart COVID-3D-SCNN: a novel method to classify X-ray images of COVID-19. *Comput. Syst. Sci. Eng.* **41**(3), 997–1008 (2022)
2. Tang, Y., Zhang, Y., Han, X., Zhang, F.-L., Lai, Y.-K., Tong, R.: 3D corrective nose reconstruction from a single image. *Comput. Vis. Media* **8**(2), 225–237 (2022)
3. Afifi, A.J., Magnusson, J., Soomro, T.A., Hellwich, O.: Pixel2point: 3D object reconstruction from a single image using CNN and initial sphere. *IEEE Access* **9** 110–121 (2021)
4. Fukuda, T., Awatsuji, Y., Xia, P., et al.: Review of three-dimensional imaging of dynamic objects by parallel phase-shifting digital holography. *Optical Eng.* **57**(6), 061613.1–061613.11 (2018)
5. Al-Masni, M.A., Kim, W.R., Kim, E.Y., Noh, Y., Kim, D.H.: 3D multi-scale residual network toward lacunar infarcts identification from MR images with minimal user intervention. *IEEE Access* **9**, 11787–11797 (2021)
6. Yzab, C., Hk, A., Hd, A., et al.: Immediate and 1-year success rate of type 2 endoleak treatment using three-dimensional image fusion guidance - ScienceDirect. *Diagn. Interv. Imaging* **101**(9), 589–598 (2020)
7. Matsumoto, R., Uechi, T., Nagasawa, Y.: Three-dimensional microstructure of frost layer measured by using X-ray μ CT. *J. Thermal Sci. Technol.* **13**(1), JTST0014–JTST0014 (2018)
8. Yushan, M., Alike, Y., Keremu, A., et al.: Precise resection of macrodactyly under assistance of three-dimensional reconstruction technology: a case report. *J. Foot Ankle Surgery Offic. publicat. Am. Coll. Foot Ankle Surg.* **59**(1), 125–127 (2020)
9. Deshpande, R.R., Bhatt, M.R., Madhavi, C.R.: Accuracy in depth recovery and 3D image synthesis from single image using multi-color filter aperture and shallow depth of field. *IEEE Access* **9**, 123528–123540 (2021)
10. Xing, Z., Yu, L., Tao, S., et al.: Treatment of severe distal humeral bone defects with three-dimensional printing technology. *Zhongguo xiu fu chong jian wai ke za zhi = Zhongguo xiufu chongjian waike zazhi = Chin. J. Reparat. Reconstruct. Surgery* **32**(12), 1534–1539 (2018)

11. Javadi, S., Dahl, M., Pettersson, M.I.: Vehicle detection in aerial images based on 3D depth maps and deep neural networks. *IEEE Access* **9**, 8381–8391 (2021)
12. Pisonero, J., et al.: A comparative study of 2D and 3D digital image correlation approaches for the characterization and numerical analysis of composite materials. *IEEE Access* **9**, 160675–160687 (2021)



Dynamic Image Monitoring Method of Hitting Point for Tennis Players in Serving Instant

Xiaopeng Ji¹(✉) and Chadi Altrjman²

¹ Physical Education Research Department, Xinjiang University, Urumqi 830046, China

² Near East University, Nicosia, Turkey

Abstract. In order to realize the dynamic monitoring ability of the hitting point for the tennis player in serving instant, thereby improving the serve quality of the tennis player and guiding the improvement of the serve action, a dynamic image monitoring method of hitting point for the tennis player in serving instant is proposed based on multi-mode registration and key contour feature point detection. Adaptive noise reduction processing is carried out on the dynamic image of hitting point for the tennis player in serving instant, the feature point and contour information of 3D hitting point's dynamic image are extracted with Harris corner detection method, and the instant hitting action of tennis player is reconstructed. The envelope detection method is used for dynamic image's multi-modal registration, and the key contour feature point detection method is used to reconstruct and monitor the dynamic image of the instant hitting point of tennis ball. The simulation results show that the proposed method is better in real time and high accuracy in detecting the key feature points of tennis players, and the performance of image recognition and reconstruction is improved.

Keywords: Tennis player · Serve · Dynamic image · Monitoring · Image reconstruction

1 Introduction

Image processing technology has been widely used to guide the movement improvement, and improve the effect of physical training. In tennis, it is necessary to accurately monitor and extract the dynamic image of the hitting point for the tennis player in serving instant, realize the 3D reconstruction of the dynamic image of hitting point for the tennis player in serving instant, and improve the service quality and the attacking ability of serving. The research on the dynamic image monitoring method of hitting point for the tennis player in serving instant has a good application value in guiding tennis training [1, 2].

The 3D reconstruction and monitoring of the dynamic image of hitting point for the tennis player in serving instant is based on the feature matching and texture rendering. Combined with corner detection and dynamic image smoothing technology, the 3D reconstruction of dynamic image of hitting point for the tennis player in serving instant is realized. Traditionally, Scale Invariant Feature Transform (SIFT) feature matching method is the main method for dynamic image monitoring of hitting point for the tennis

player in serving instant. In addition, the traditional methods include edge contour feature decomposition method [3, 4], support vector machine (SVM) method and sparse point reconstruction method, etc. The real-time monitoring of dynamic image is not good, and the output quality of dynamic image is not high [5]. In this paper, a dynamic image monitoring method based on multi-mode registration and key contour feature point detection is proposed. The dynamic image of hitting point for the tennis player in serving instant is processed by adaptive noise reduction, and the dynamic image is registered by envelope detection method. The key contour feature point detection method is used to reconstruct and monitor the dynamic image of hitting point for the tennis player in serving instant. Finally, the simulation experiment is carried out. This paper shows the superior performance of this method in realizing the dynamic image monitoring of hitting point for the tennis player in serving instant.

2 3D Template Matching and Dynamic Image Denoising of Tennis Players' Instant Batting Point

2.1 3D Template Matching of Dynamic Images of Tennis Players' Instant Batting Points

In order to realize the high-resolution monitoring of the dynamic image of hitting point for the tennis player in serving instant, the characteristic analysis model of the dynamic image is constructed. The 3D template matching method is used to detect the edge correlation of the dynamic image of hitting point for the tennis player in serving instant [6]. The pixel sequence distribution matrix of the dynamic image is:

$$D = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (1)$$

The dynamic scene feature matching method is used to extract the hitting point for the tennis player in serving instant, the edge pixel feature components of the 3D multi-modal dynamic image are obtained:

$$E(\mathbf{T}_n) = \frac{1}{N} \sum_{i=0}^{N-1} \|p_i - \mathbf{T}_n(g'_i)\|^2 + \tau \cdot \Phi(\mathbf{T}_n) \quad (2)$$

According to the distribution of the surface texture structure of the dynamic image, the inverse weighting of the dynamic image of hitting point for the tennis player in serving instant is obtained. The weighted function $f(g_i)$ is calculated as:

$$f(g_i) = c_1 \tilde{\lambda}_i \sum_{j=0}^{N_{np}} \frac{\rho_j \tilde{v}_{ij}}{|\tilde{v}_{ij}|^{\sigma_1} + \varepsilon} / \sum_{j=0}^{N_{np}} \frac{\rho_j}{|\tilde{v}_{ij}|^{\sigma_1} + \varepsilon} \quad (3)$$

The background difference component of 3D dynamic image of hitting point for the tennis player in serving instant is obtained, and the RGB decomposition formula of the

smooth region of the dynamic image is expressed as follows:

$$\begin{cases} f(\mathbf{G}_n) = a_1 + a_2x + a_3y + a_4z + \sum_{i=0}^n \gamma_i U(g'_i, p_i) \\ g(\mathbf{G}_n) = b_1 + b_2x + b_3y + b_4z + \sum_{i=0}^n \theta_i U(g'_i, p_i) \\ h(\mathbf{G}_n) = c_1 + c_2x + c_3y + c_4z + \sum_{i=0}^n \omega_i U(g'_i, p_i) \end{cases} \quad (4)$$

Where

$$\Phi(\mathbf{T}_n) = \gamma^T \mathbf{H}\gamma + \theta^T \mathbf{H}\theta + \omega^T \mathbf{H}\omega \quad (5)$$

According to the above model, the 3D template matching method is used to detect the edge correlation of the dynamic image of hitting point for the tennis player in serving instant, and the 3D template matching of the dynamic image is realized [7].

2.2 Image Adaptive Noise Reduction

In this paper, the dynamic image of hitting point for the tennis player in serving instant is processed by adaptive noise reduction, and the feature points and contour information of 3D dynamic image of hitting point for the tennis player in serving instant are extracted by using Harris corner detection method. The texture feature segmentation of the monitoring surface point scanning method of the dynamic image is carried out, and the noise distribution function of the image is obtained as follows:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (6)$$

Where

$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y) \quad (7)$$

In the form, $I(x, y)$ represents the dynamic image of hitting point for the tennis player in serving instant is divided into blocks at (x, y) , and the resolution coefficient $L(x, y, \sigma)$ of the dynamic image of hitting point for the tennis player in serving instant is expressed as the pixel value $G(x, y, \sigma)$. The formula is as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (8)$$

By using the correlation detection method and adaptive noise reduction technique, the gray pixel value $E(d(x, y))$ of the dynamic image of hitting point for the tennis player in serving instant is obtained. The sparse degree control function of the 3D monitoring of hitting point for the tennis player in serving instant is expressed as follows:

$$F_d - \frac{d}{dx} F_{d_x} - \frac{d}{dy} F_{d_y} = 0 \quad (9)$$

Where, F_d denotes the edge scale of the dynamic image of hitting point for the tennis player in serving instant, and F_d denotes the noise distance of the mesh surface of the dynamic image [8].

3 Dynamic Image Monitoring Optimization of Tennis Players' Instant Batting Point

Combined with the Harris corner detection method, the feature points and contour information of the 3D dynamic image are extracted [9], the hitting action for the tennis player in serving instant is reconstructed, and the original dynamic image of hitting point for the tennis player in serving instant is filtered adaptively. The filter function is:

$$\mathbf{G}_{new} = (1 + \mu T)(1 + \lambda T)\mathbf{G}_{old} \quad (10)$$

$$T(g_i) = \frac{1}{\sum_k \omega_k} \cdot \sum_k \omega_k (g_k - g_i) \quad (11)$$

Based on the Harris corner detection method, the feature points and contour information of hitting point for the tennis player in serving instant are extracted [10], and the information components of the dynamic action are obtained as follows:

$$L(a, b_m) = \log\left(\frac{|V||V_m \cap V_n|}{|V_m||V_n|}\right) \quad (12)$$

Multi-mode registration of the surface texture features of 3D dynamic image is carried out, and the output is:

$$L(a, b_m) = \sum_{V_m \in P^{res}} \sum_{V_n \in P^{true}} \frac{|V_m \cap V_n|}{|V|} \log\left(\frac{|V||V_m \cap V_n|}{|V_m||V_n|}\right) \quad (13)$$

The average pixel set of the dynamic image of hitting point for the tennis player in serving instant is:

$$\bar{x}_T = \frac{1}{T} \sum_{i=1}^T x_i \quad (14)$$

$x_1, x_2, x_3 \dots x_T$ are the template matching set of each subblock, T is the pixel distribution density. The statistical shape model of the dynamic image of hitting point for the tennis player in serving instant is established, and the edge pixel set of 3D monitoring image is:

$$F = \tilde{p}(x, y) = p(x, y) \left(\frac{v(x)}{v(y)}\right)^{1/2} \quad (15)$$

Where

$$p(x, y) = \frac{k(x, y)}{v(x)}, \quad v(x) = \sum_y k(x, y) \quad (16)$$

The asynchronous iterative updating method is used, the adaptive registration of the dynamic image and the high resolution monitoring of the dynamic image are carried out. Registration is used to reconstruct and monitor the dynamic image of hitting point for the tennis player in serving instant.

4 Simulation Experiment and Result Analysis

In order to test the application performance of the proposed method in the realizing high-resolution monitoring image of hitting point for the tennis player in serving instant, the simulation experiment is carried out, and the experiment is designed by Matlab. The number of sparse points sampled from dynamic image of hitting point for the tennis player in serving instant is 100, the regularization parameter of image matching is 0.23, the smoothing operator of dynamic image monitoring is 1.24, and the fuzzy kernel coefficient is 0.4. According to the above simulation environment and parameter setting, the number of sparse points is 100, and the regularization parameter of image matching is 0.26. Image monitoring simulation is taken, and the original image is shown in Fig. 1.



Fig. 1. Dynamic image of instant batting point of tennis player serving

The image collected in Fig. 1 is taken as the research object, the feature extraction and dynamic monitoring image of hitting point for the tennis player in serving instant are performed, and the result of the dynamic monitoring is shown in Fig. 2.

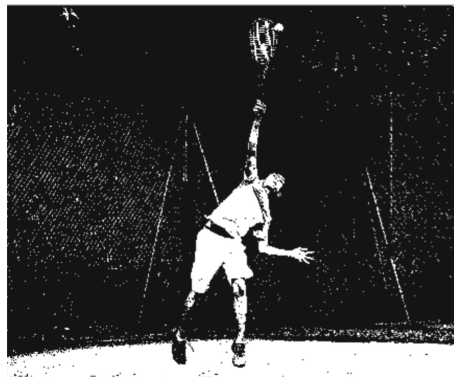


Fig. 2. Monitoring results of dynamic characteristics of tennis serve

Figure 2 shows that the dynamic feature monitoring and the feature extraction of hitting point for the tennis player in serving instant can be realized effectively by using the proposed method. The performance of feature registration is tested and the result is shown in Fig. 3. Figure 3 shows that the proposed method has better feature registration for dynamic image detection of hitting point for the tennis player in serving instant.

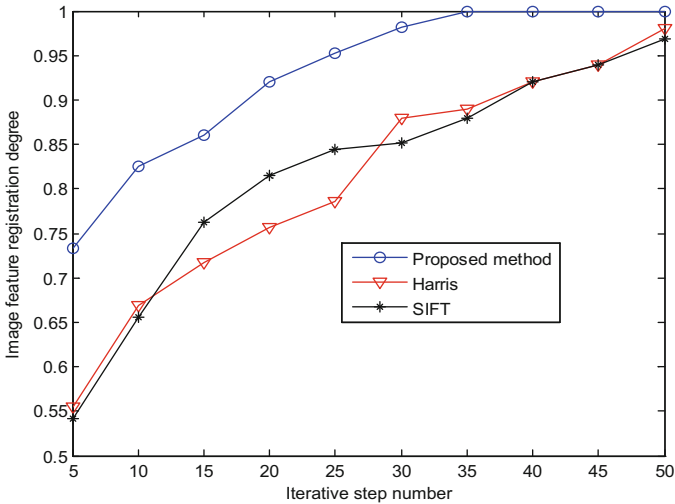


Fig. 3. Comparison of feature registration in image monitoring

5 Conclusions

In this paper, a dynamic image monitoring method of hitting point for the tennis player in serving instant is proposed based on multi-mode registration and key contour feature point detection. Adaptive noise reduction processing is carried out on the dynamic image of hitting point for the tennis player in serving instant, the feature point and contour information of 3D dynamic image is extracted by Harris corner detection method, and the instant hitting action of tennis player is reconstructed. The envelope detection method is used for dynamic image multi-modal registration, and the key contour feature point detection method is used to reconstruct and monitor the dynamic image of hitting point for the tennis player in serving instant. The simulation results show that this method is better in real time and high accuracy in detecting the hitting point for the tennis player in serving instant, and the performance of image recognition and reconstruction is improved. This method has good application value in tennis training and instruction.

Acknowledgement. This work was supported by “National Social Science Foundation of China (NO. BLA200218)”, Research on the construction of physical education curriculum objectives in Xinjiang multi-ethnic areas from the perspective of national identity.

References

1. Giles, B., Reid, M.: Applying the brakes in tennis: how entry speed affects the movement and hitting kinematics of professional tennis players. *J. Sports Sci.* **39**(3), 1–8 (2020)
2. Herbaut, A., et al.: The influence of shoe drop on the kinematics and kinetics of children tennis players. *Eur. J. Sport Sci.* **16**(8), 1–9 (2016)
3. Thibault, B., Morio, C., Millet, G.Y., Rossi, J.: Influence of shoe drop on running kinematics and kinetics in female runners. *Eur. J. Sport Sci.* **19**(16), 1–8 (2019)
4. Landlinger, J., Lindinger, S.J., Stöggl, T., Wagner, H., Müller, E.: Kinematic differences of elite and high-performance tennis players in the cross court and down the line forehand. *Sports Biomech.* **9**(4), 280–295 (2010)
5. Yam, J.W., Pan, J.W., Kong, P.W.: Measuring upper limb kinematics of forehand and backhand topspin drives with IMU sensors in wheelchair and able-bodied table tennis players. *Sensors (Basel)* **21**(24), 8303 (2021)
6. Siener, M., Faber, I., Hohmann, A.: Prognostic validity of statistical prediction methods used for talent identification in youth tennis players based on motor abilities. *Appl. Sci.* **11**(15), 7051 (2021)
7. Bagiwa, M.A., Wahab, A.W.A., Idris, M.Y.I., Khan, S.: Digital video inpainting detection using correlation of hessian matrix. *Malays. J. Comput. Sci.* **29**(3), 179–195 (2016)
8. Khasawneh, M.A.: The impact of using an infographic-based training program on improving visual motor memory among EFL students. *Indian J. Sci. Technol.* **13**(10), 405–416 (2020)
9. Zhao, Q., Tan, H.R., Wang, X.Y., Teng, K.D.: Research progress of the electrode for electroencephalogram acquisition. *Sci. Technol. Eng.* **21**(15), 6097–6104 (2021)
10. Zhou, Y.M., Li, G.Y.: Image smoothing beautification processing algorithm in multi-illumination color difference. *Comput. Sci.* **43**(10), 287–291 (2016)



Cardiovascular Monitoring System Design Based on Medical Imaging Technology and Artificial Intelligence Algorithm

Ruiliang Ling¹(✉) and Amar Jain²

¹ Happ Your Mart in Shoukai Fumao Building, Beijing 100079, China

² Madhyanchal Professional University, Bhopal, India

Abstract. With the rapid development of medical imaging technology and artificial intelligence algorithms, cardiovascular monitoring systems have attracted more and more attention. The rapid progress and development of these cardiovascular monitoring technologies have provided a lot of help for the prevention and treatment of cardiovascular and cerebrovascular diseases. Research on cardiovascular monitoring systems has become an important topic. Therefore, this article designs a cardiovascular monitoring system based on medical imaging technology and artificial intelligence algorithms. This article introduces the ECG signal acquisition module and signal preprocessing module of the system in detail, and uses the wavelet decomposition algorithm in medical imaging technology to optimize the processing of the ECG signal. Aiming at the cardiovascular monitoring system designed in this article, this article has carried out targeted experiments. Experimental data shows that the standard deviation of systolic blood pressure measured by the sample machine is within 2.5, and the standard deviation of diastolic blood pressure is within 3.5. This shows that the system has high measurement accuracy.

Keywords: Cardiovascular monitoring · Pulse wave signal · Artificial intelligence algorithm · Medical imaging technology

1 Introduction

With the rapid development of science and technology and people's requirements for the advancement of medical standards, intelligent medical care based on cardiovascular monitoring systems has gradually entered people's field of vision [1, 2]. Using mature medical imaging technology and artificial intelligence algorithms to monitor ECG signals in real time plays an important role in preventing the occurrence of cardiovascular diseases [3, 4]. At the same time, related data collection and preprocessing technology has promoted the development of smart medicine [5, 6].

Regarding the research of cardiovascular monitoring, many scholars have conducted multi-angle investigations. For example, Rincon JA designed a cardiovascular patient monitoring system based on the Internet of Things and fog computing [7]; Kapoor A developed the heart of cancer survivors. Experimental research on the monitoring of risk factors for vascular diseases [8]; Fujiwara T conducted a research on the risk of

cardiovascular events related to home blood pressure monitoring and occult nocturnal hypertension [9]. It can be seen that this article combines medical imaging technology, artificial intelligence algorithms and cardiovascular monitoring for research, which has important practical significance.

This article first systematically introduces the ECG signal acquisition module and the signal preprocessing module, including the light source drive module, the receiving module, and the ECG signal amplifying circuit. Then the wavelet decomposition algorithm is used to optimize the processing of the ECG signal data, and finally through experiments, the accuracy of the system in measuring cardiovascular data is verified.

2 Cardiovascular Monitoring System Design Based on Medical Imaging Technology and Artificial Intelligence Algorithms

2.1 ECG Signal Acquisition Module

(1) Light source drive module

The ECG signal has a low frequency and is very susceptible to noise interference. Therefore, a receiving module with high sensitivity, good linearity, and fast response should be selected. The dual-wavelength common anode photodiode used in the system, the model is 2CU10s, has the advantages of low dark current, high sensitivity, fast response, little influence by external factors such as temperature, and reasonable price.

The light source drive module makes the common anode LED emit light according to a specific time sequence, and must ensure that the incident light intensity is a constant value. If the incident light intensity is not constant, the received transmitted light cannot reflect the pulse wave signal [10, 11].

(2) Photoelectric receiving module

This system uses a transmissive photoelectric sensor. A 660 nm red LED and a 940 nm infrared LED form a common anode light emitting diode. The photodiode at the receiving end converts the transmitted light signal into a current signal, then converts the current signal into a voltage signal, and finally stores the collected data for the next step of preprocessing.

(3) Amplifying circuit of ECG signal

The ECG signal is directly taken from the body surface through the electrode sheet. The electrode sheet is a silver-silver chloride electrode, because the half-cell electromotive force of this electrode is much more stable than other material electrodes, and the polarization potential is also very small. ECG signal frequency is low (0.05–100 Hz), relatively weak (0.05–5 mV), signal source impedance is high (a few thousand ohms to tens of thousand ohms) and often accompanied by strong background noise, so it is necessary to design the front amplifying circuit. The system selects the AD620 chip produced by Analog Devices as the main chip of the preamplifier. It has many excellent characteristics, such as low noise, low power consumption, high input impedance and high mode rejection ratio, etc., and it has great effect on very weak input signals.

Since the amplitude of the ECG signal collected by the entire system generally does not exceed 5 mV, the input voltage of the A/D converter is required to be 0–3.3V, so the entire collection circuit is required to be increased by 500 times. The system sets

the gain of the preamplifier to 10 times, and the gain of the main amplifier circuit to 50 times.

2.2 Signal Preprocessing Module

Preprocessing the collected signal is to filter out the interference caused by noise. The main process is: first select the appropriate wavelet base for wavelet decomposition to obtain the wavelet coefficients of each scale; then obtain the threshold in the denoising process, perform threshold processing, and obtain the estimated coefficient; finally, perform wavelet reconstruction according to the estimated coefficient to obtain the denoising after the signal.

As a scale and multi-channel analysis method, wavelet decomposition is very suitable for multi-scale extraction test of medical image edge information [12].

Here, assuming that the integral of $g(x, y)$ on the plane is 1, and converges to 0 at infinity at xy , then $g(x, y)$ is a 2-D smooth function, and then define wavelet functions $w_1(x, y)$ and $w_2(x, y)$, calculate The method is shown in formulas (1) and (2):

$$w_1(x, y) = \frac{\partial g(x, y)}{\partial x} \tag{1}$$

$$w_2(x, y) = \frac{\partial g(x, y)}{\partial y} \tag{2}$$

The two components of image $f(x,y)$ in wavelet transform are defined in s scale as shown in formulas (3) and (4):

$$w_{1,s}f(x, y) = f * w_{1,s}(x, y) \tag{3}$$

$$w_{2,s}f(x, y) = f * w_{2,s}(x, y) \tag{4}$$

Perform wavelet decomposition to get formula (5):

$$\begin{pmatrix} w_{1,2}f(x, y) \\ w_{2,2}f(x, y) \end{pmatrix} = \begin{pmatrix} \partial/\partial x(f * g_2)(x, y) \\ \frac{\partial}{\partial y(f * g_2)(x, y)} \end{pmatrix} \tag{5}$$

From Eq. (5), it can be seen that the point of maximum value on $f * g_2(x,y)$ is the size of the two angles. The signal-to-noise ratio (SNR) and mean square error (MSE) are selected to determine the filtering effect, and their mathematical models are formula (6) and formula (7) respectively:

$$SNR = 10 \log_{10} \sum_{i=1}^N \frac{y_i^2}{(x_i - y_i)^2} \tag{6}$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2 \tag{7}$$

In the formula, it is the original signal and the signal after denoising. And, the larger the SNR, the smaller the MSE, indicating that the signal denoising effect is better. After many comparison experiments, the system finally adopts sym8 wavelet to decompose the collected ECG signal and pulse signal by wavelet, and then carry out wavelet reconstruction.

2.3 Design of Playback Display Module

This module allows users to view the data collected by the system, including operations such as displaying waveforms and turning pages. The system uses double-buffer drawing technology to draw pulse waveforms and ECG waveforms. The basic principle is to use a DC compatible with the current window to set the waveform in memory in advance, and then transplant the pre-drawn waveform to the current window. Because it does not draw images directly on the application projection screen, there will be no screen flickering.

3 Experimental Design of Cardiovascular Monitoring System Based on Medical Imaging Technology and Artificial Intelligence Algorithms

3.1 Purpose of the Experiment

Verify the accurate performance of the system prototype in measuring blood pressure.

3.2 Subjects

This experiment selected 10 healthy volunteers, 5 males and 5 males. None of the subjects had high blood pressure, diabetes or other cardiovascular diseases.

3.3 Experimental Equipment

Under room temperature conditions, the experiment uses more clinically used mercury sphygmomanometers to obtain the systolic and diastolic blood pressure of the human body. The main performance indicators are as follows:

- (1) Standard double scale display: two readings of kPa and mmHg.
- (2) Measuring range: 0–40 kPa (0–300 mmHg).
- (3) Minimum graduation value: 0.5 kPa (2 mmHg).
- (4) The basic error range: ± 0.5 kPa (± 3.75 mmHg).

In addition, the experimental equipment includes a prototype of the system, a mercury sphygmomanometer, a disposable ECG electrode, medical alcohol and cotton swabs.

3.4 Experimental Procedure

In order to collect complete and accurate ECG and pulse data, the experiment conducted 6 data sampling for each tester. The test time is 8:30–11:30 in the morning and 2:00–5:00 in the afternoon, each time period Test 3 times each.

The specific experiment process is as follows:

- (1) Dip a cotton swab with a small amount of medical alcohol, and wipe the position of the three-lead tester's ECG measurement and the position of the fingertip to be tested;

- (2) Put the electrode chip on the right arm of the test subject, the fingertip pulse sensor on the left glove, and the left arm on the table;
- (3) Place a stethoscope on the femoral artery of the right arm of the tester, and wear the cuff of a mercury sphygmomanometer for the tester to fix the stethoscope, and place the right arm on the workbench;
- (4) The person to be tested needs to maintain a normal sitting posture, check whether each device is worn accurately and firmly, and start the system prototype after the preparation work is completed;
- (5) One minute after the system prototype was started, the test staff began to inflate the cuff of the mercury sphygmomanometer at a constant speed and perform blood pressure measurement. The measurement time for each tester was approximately 4.5 min;
- (6) After the measurement, record the diastolic and systolic blood pressure data measured by the mercury sphygmomanometer.

3.5 Performance Analysis

The ECG and pulse data collected by the system prototype will be stored in a computer file, and the corresponding file name will correspond to the blood pressure value measured by the mercury sphygmomanometer, and the data will be processed using MATLAB tools. A total of 10 sample data were collected in this experiment, each sample collected 6 sets of data, a total of 60 sets of data (systolic blood pressure SBP-diastolic blood pressure DBP). Use this data as the standard blood pressure data to judge the accuracy of the blood pressure values collected by the prototype.

Due to the large amount of data, the 160 sets of data collected by the system prototype are first calculated according to the corresponding tester's 1–10 number, and the average values of each person's 6 groups of systolic and diastolic blood pressure are calculated as the collection results of the system prototype; the same process is performed, and the mean value is used as the final standard blood pressure data.

4 Experimental Analysis of Cardiovascular Monitoring System Based on Medical Imaging Technology and Artificial Intelligence Algorithms

4.1 Statistics of Absolute Measurement Error Between Prototype and Mercury Sphygmomanometer

When collecting sample data on the system prototype, the mercury sphygmomanometer collects the same number of sets of blood pressure data (systolic blood pressure-diastolic blood pressure) at the same time. The absolute value error collected by the two devices is shown in Table 1: among the systolic blood pressure SBP values, the maximum absolute value error collected by the two devices is 4; among the diastolic blood pressure DBP values, the maximum absolute value error collected by the two devices is 5.

Table 1. Statistic absolute error of absolute measurement error between prototype and mercury sphygmomanometer (mmHg)

Sample number	SBP	DBP
1	1	4
2	2	5
3	2	4
4	1	3
5	3	3.5
6	4	4
7	2	3
8	1.5	4
9	1	4
10	1	3

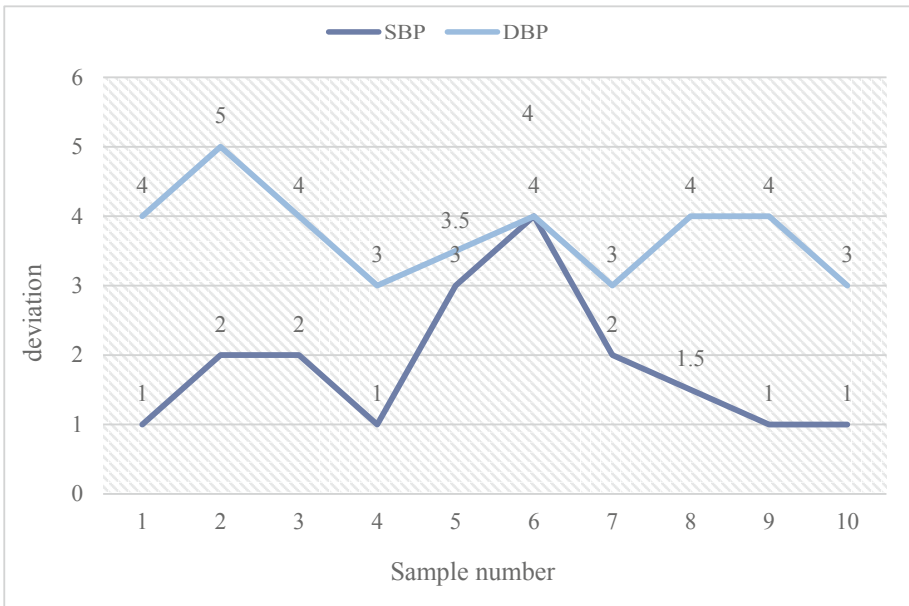


Fig. 1. Statistic absolute error of absolute measurement error between prototype and mercury sphygmomanometer (mmHg)

Observing Fig. 1 can draw the following conclusions: referring to the international AAMI standard, the absolute value of the average measurement error of the sphygmomanometer does not exceed 5 mmHg, it is deemed that the measurement meets the standard. Regardless of whether it is systolic blood pressure or diastolic blood pressure,

the absolute value of the error between the data collected by the system prototype and the standard data of the mercury sphygmomanometer is within 5mmHg. Therefore, it can be considered that the cardiovascular monitoring system designed in this paper meets the requirements of AAMI and has high accuracy.

4.2 Standard Deviation Statistics of Sample Measurement Results

In addition, this article also counts the standard deviations of the systolic and diastolic blood pressures of the 8 test samples, as shown in Table 2. Among the 10 testers, the maximum standard deviation of the prototype systolic blood pressure is 2.65 and the minimum is 1.34; diastolic blood pressure the maximum standard deviation is 3.31 and the minimum is 1.42.

Table 2. Standard deviation of sample measurement results

Sample number	SBP Mean (mmHg)	SBP standard deviation	DBP Mean(/mmHg)	DBP standard deviation
1	118	1.45	79	1.28
2	127	2.35	80	3.31
3	127	1.34	81	3.09
4	123	2.03	82	2.45
5	117	1.34	79	1.42
6	108	2.04	78	2.54
7	124	1.50	79	2.79
8	121	2.01	82	1.91
9	128	2.03	80	1.91
10	122	2.14	81	1.24

It can be seen from Fig. 2 that the error between the systolic blood pressure measured by the system prototype and the data measured by the mercury sphygmomanometer is relatively small. The standard deviation of the systolic blood pressure measured by the system prototype is within 2.5, and the standard deviation of diastolic blood pressure is within 3.5. The volatility is smaller, and it can basically be determined that the actual measurement of the system prototype has better repeatability.

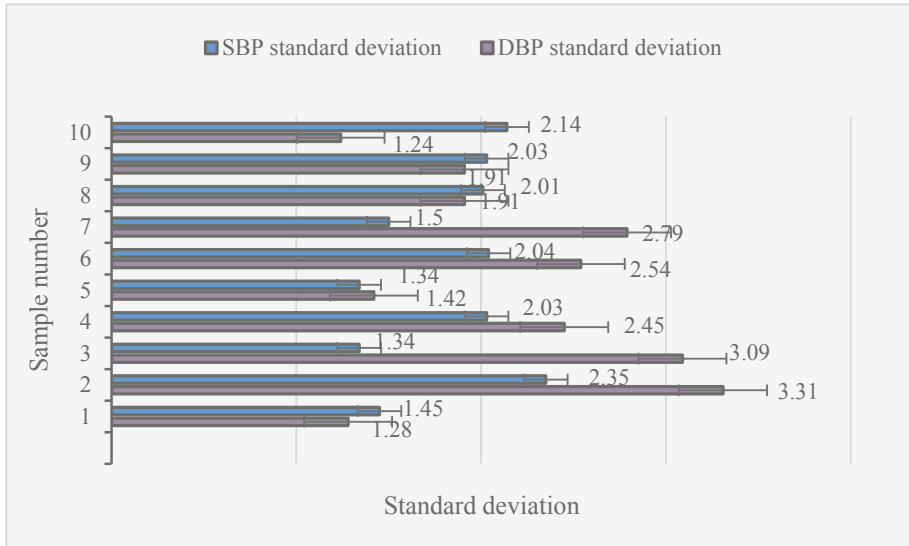


Fig. 2. Standard deviation of sample measurement results

5 Conclusions

As people pay more and more attention to healthy life, real-time monitoring of human cardiovascular system has received more and more attention. The development of medical imaging technology and artificial intelligence algorithms provides technical support for the cardiovascular monitoring system. This paper designs the system from the ECG signal acquisition module and signal preprocessing module, and uses wavelet decomposition algorithm to optimize the processing of the ECG signal data. This paper verifies the accuracy of the system measurement by comparing the systolic and diastolic blood pressure data collected by the mercury sphygmomanometer and the system prototype.

References

1. Xu, H., Liu, J., Zhang, J., et al.: Flexible organic/inorganic hybrid near-infrared photoplethysmogram sensor for cardiovascular monitoring. *Adv. Mater.* **29**(31), 1700975.1–1700975.6 (2017)
2. Surrel, G., Teijeiro, T., Chevrier, M., et al.: Event-triggered sensing for high-quality and low-power cardiovascular monitoring systems. *IEEE Des. Test*, **37**(5), 85–93 (2019)
3. Lenz, A., Petersen, J., Riedel, C., et al.: 4D flow cardiovascular magnetic resonance for monitoring of aortic valve repair in bicuspid aortic valve disease. *J. Cardiovasc. Magn. Reson.* **22**(1), 29 (2020)
4. Agarwala, R., Wang, P., Bishop, H.L., et al.: A 0.6V 785-nW multimodal sensor interface IC for ozone pollutant sensing and correlated cardiovascular disease monitoring. *IEEE J. Solid-State Circ.* **56**(4), 1058–1070 (2021)
5. Luo, J., Zhen, J., Zhou, P., et al.: An iPPG-based device for pervasive monitoring of multi-dimensional cardiovascular hemodynamics. *Sensors* **21**(3), 872 (2021)

6. Filippo, P.D., Malanchini, G., Ferrari, P.: Remote monitoring of cardiovascular implantable electronic devices: a valuable tool for a better care. *Kardiol. Pol.* **78**(11), 1086–1087 (2020)
7. Rincon, J.A., Guerra, S., Carrascosa, C., et al.: An IoT and fog computing-based monitoring system for cardiovascular patients with automatic ECG classification using deep neural networks. *Sensors* **20**(24), 7353 (2020)
8. Kapoor, A., Prakash, V., Sekhar, M., et al.: Monitoring risk factors of cardiovascular disease in cancer survivors. *Clin. Med. (Lond.)* **17**(4), 293–297 (2017)
9. Fujiwara, T., Hoshida, S., Kanegae, H., et al.: Cardiovascular event risks associated with masked nocturnal hypertension defined by home blood pressure monitoring in the J-HOP nocturnal blood pressure study. *Hypertension* **76**(1), 259–266 (2020)
10. Pena, M.J., Stenvinkel, P., Kretzler, M., et al.: Strategies to improve monitoring disease progression, assessing cardiovascular risk, and defining prognostic biomarkers in chronic kidney disease. *Kidney Int. Supp.* **7**(2), 107–113 (2017)
11. Muniyappa, A., Judson, G.L., Shen, H., et al.: B-PO04-048 patient characteristics associated with adherence to remote monitoring of cardiovascular implantable electronic devices. *Heart Rhythm* **18**(8), S298 (2021)
12. Cwh, A., Ssh, B.: Role of cardiovascular biomarkers in the risk stratification, monitoring, and management of patients with cancer. *Cardiol. Clin.* **37**(4), 505–523 (2019)



Electrical Control System Based on Machine Algorithm and Adaptive Fuzzy Algorithm

Lu Zhou^{1(✉)} and Yu Cui²

¹ School of Electrical and Information Engineering, Liaoning Institute of Science and Technology, Benxi 117004, Liaoning, China

² Siemens Ltd., China, Beijing 110000, China

Abstract. With the rapid development of industrial technology, it has become more and more proficient in the application of computer networks and modern communication technology. For the electrical control system, the most important thing is to set the parameters of the system. As a global search algorithm, the fuzzy algorithm is widely used in practical applications. For this reason, this article proposes an electrical control system based on machine algorithm and adaptive fuzzy algorithm, which is intended to have a deeper understanding of electrical control system. This article mainly uses the experimental method and the comparative method to study the algorithm of the electrical control system. The experimental results show that when the speed is 300r/min, the steady-state error of the three motors is about 5%, and the gap is relatively small. The adaptive fuzzy algorithm has a good control effect in the electrical control system.

Keywords: Machine algorithm · Adaptive fuzzy algorithm · Electrical control · System design

1 Introduction

The power system is a complex, highly nonlinear and time-varying, and it also requires good stability and relatively stable operation. Adaptive fuzzy algorithm is an important branch of intelligence, and can solve the problem of fuzzy system model fuzziness. How to ensure the normal and stable operation of the system is an important issue. The adaptive fuzzy algorithm can adjust and modify the parameters of the inverter. This has very good guiding significance for our future practice.

There are many research theories on electrical control systems based on machine algorithms and adaptive fuzzy algorithms. For example, according to R. R. Soria, PMSMs offer better performance and longer lifetimes, while slightly increasing the cost and complexity of the system. In a vector-controlled PMSM drive system, the resolution of the rotor position sensor plays a crucial role in the overall performance of the system [1]. Yuri S developed a stochastic prediction method based on the generation of optimal entropy prediction trajectory sets. This program is used for stochastic forecasting of daily electrical loads in regional power systems [2]. JiL. N. Das believes that the micro-

and macro-network control functions of utilities and multi-unit households have facilities to measure synchronously and asynchronously emitted phase current pulses [3]. Therefore, the electrical control system of this paper has carried out the research of machine algorithm and adaptive fuzzy algorithm.

This article first studies genetic algorithm and its adaptive design, and describes these two concepts. Secondly, the fuzzy PID control algorithm is deeply studied. Then the electrical control system is analyzed and designed. Finally, relevant experiments are carried out on the electrical control system, and the data results are obtained.

2 Electrical Control System Based on Machine Algorithm and Adaptive Fuzzy Algorithm

2.1 Genetic Algorithm and Its Adaptive Design

When a genetic algorithm solves a specific problem, it must establish a connection between the target problem and the chromosome of the algorithm. This connection is established through coding. The fitness function has different shapes according to specific problems, and its accuracy directly affects the convergence speed of the genetic algorithm, and the optimization results between different fitness functions are also different. Since the calculation of fitness is cumbersome, the evaluation of fitness has become an important factor limiting the performance of genetic algorithms [4, 5].

Genetic algorithms use fitness to measure individual optimization. Once the individual code string is decoded, the individual phenotype can be obtained. Calculate the individual's objective function value from the phenotype. According to the needs of actual problems, the target value is converted into fitness [6, 7].

The operation of selecting outstanding individuals from the group and eliminating inferior individuals is called selection operation. The selection process is based on the evaluation of the suitability of individuals in the group and aims to pass outstanding individuals directly to the next generation. Because of its global search capability, the crossover operator plays a central role in genetic algorithms. The hybridization operation simulates the phenomenon of sexual reproduction in the process of biological evolution by cross-combining two chromosomes to produce new outstanding individuals, thereby greatly improving the research capabilities of genetic algorithms. Mutation operation is a necessary tool for genetic algorithm to maintain population diversity [8, 9].

The termination condition of the genetic algorithm can also be a specific criterion, as long as it is determined that the population has matured and there is no longer any evolutionary trend, the operation of the algorithm can be terminated. There are two commonly used results: the average fitness difference of individuals in successive generations is below a certain minimum threshold, and the fitness variance of all individuals in the group is below a certain minimum threshold [10, 11].

2.2 Fuzzy PID Control Algorithm

PID controller has simple structure, good stability, reliable work and convenient adjustment, and it has become one of the most important industrial control technologies.

When the structure and parameters of the controlled object cannot be fully grasped or an accurate mathematical model cannot be obtained, and other control theory techniques are difficult to apply, the structure and parameters of the system controller should be determined through experience and experience. It is most convenient to debug Site. PID control technology is the most convenient [12, 13]. PID controller is a kind of linear controller, which is based on the control deviation between the given value and the actual output value:

$$\text{error}(s) = R(s) - Y(s) \tag{1}$$

The control law of PID is:

$$H(q) = \frac{P(q)}{F(q)} = l_k \left(1 + \frac{1}{S_1 q} + S_d q \right) \tag{2}$$

In the formula, l_k is the proportional coefficient, S_1 is the integral time constant, and S_d is the derivative time constant. PID control is one of the earliest control strategies developed. With the introduction of microcomputer processing technology and modern control theory, many new PID controllers have appeared.

The essence of fuzzy control is to transform the professional knowledge and experience of skilled operators in related fields into fuzzy language rules, and realize the control of complex systems through fuzzy reasoning and fuzzy decision-making. The difference between a fuzzy control system and a traditional automatic control system is that its core is a knowledgeable intelligent fuzzy controller. The structure of the fuzzy control system has the structure of a conventional computer control system, as shown in Fig. 1.

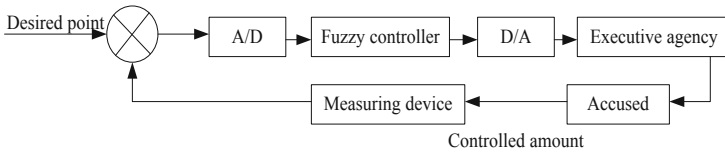


Fig. 1. Fuzzy control system composition structure

As shown in the figure, the fuzzy control system generally consists of five parts: fuzzy controller, input/output interface, actuator, controlled object and measuring device. An adaptive system is a system with a certain degree of adaptability. It can detect changes in environmental conditions and automatically correct regular actions to make the system achieve optimal or sub-optimal control effects.

2.3 Design of Electrical Control System

In the construction of the electrical control system, it is mainly composed of the following parts. The output voltage directly affects the normal operation of the system. Therefore, it is necessary to amplify and transform the input signal to make the entire device

work stably. The control function module can execute various control commands and instructions, receive data and transmit the analysis results to the actuator to perform corresponding driving tasks. At the same time, it can judge whether the control board should start or stop the power supply and various circuits according to the received information, switches and other actions. In relay control, the main system consists of switches, voltage converters and power devices. Among them, the switch generates logical operations through the flow of contacts to perform various operations. The load signal can be directly input to the single-chip microcomputer for processing, and then sent to a stepper motor or other load operation to achieve the purpose of speed control. The pulse amplitude can be used to determine whether the engine is starting or stopping and to issue control commands. When designing control, various switch types can be used to control electrical equipment such as motors, bridge rectifiers, and filter capacitors. Played a very important role in the entire design process is the main control part of the microcontroller and the writing and debugging of the corresponding software program. The auxiliary control unit must provide the output terminals of the voltage and current signals required by each functional component to coordinate the working relationship between the various sub-organizations and achieve their basic functional requirements and goals. PWM voltage control loop: can form a group of switches. It can also consist of a set of digital or analog inputs from the central processing unit and a part of the unit circuit for processing operations. It is also possible to construct functional modules that can generate square wave pulses to achieve the level of power quality control.) PFM control: It can pulse-encode the input voltage signal and then send it to the next processor. Its working principle is to use a timer to achieve period and frequency synchronization.

The electrical control system mainly has the following working modes: According to the control mode, it is divided into open loop and closed loop. The open-loop feature is that it can continuously adjust parameters such as motor stop, current, and temperature in real time to ensure that the system can safely and reliably meet various functional requirements during operation.

Subdivided by operation mode, electrical control can be divided into centralized and decentralized. In the case of decentralized control, if the distributed parameterization method is used for speed control or speed control, it is called a centralized system. On the contrary, it is a distributed control system. The control system that uses centralized control to achieve small-scale, multi-destination continuous automatic adjustment without adjusting the speed and RPM is called a distributed system (DCS). According to the different control objects, it can be divided into two types: constant speed, dynamic or constant speed. Usually the output of the current regulator remains stable. But sometimes it is also possible to use DC or oscillating voltage regulation to achieve higher speed requirements and use equivalent resistance.

The circuit is the "brain" of the entire system. It assigns different control commands to different components at different times and stages. Each functional unit is logically and physically separated into a whole, creating a unified action, suitable for collaborative processing and meeting operational requirements. The design of the electrical control system controls electrical equipment through various methods to ensure the normal operation of the entire system. The design process must consider possible errors and uncertainties.

(1) Electrical control system design

In the design process, we must choose the microcontroller wisely according to the system requirements and carry out necessary and appropriate tests. The control circuit in the electrical control system design is mainly composed of relays, digital electronic components, and single-chip microcomputers. We can use PLC as a system controller and use its logic function to process the received current signals quickly and accurately. Before starting the system, the thermal stability must be calibrated and the cold temperature must be adjusted. Start the motor when the power supply voltage is lower than the set value, and vice versa. Functional modules such as motor stop rotation, power disconnection or load current cut-off, and thermal protection circuit to ensure reliable working environment conditions for all equipment.

(2) Basic requirements for electrical control design

Based on the principle of complete functions and complete control equipment, the system can work stably and reliably. This is to ensure that the controlled object is in a good state and will not produce harmonics; if the disturbance is large or the strong signal is quickly transformed into a disturbance signal and the transient response is slow, if there is vibration or even instability, the influence should be eliminated as soon as possible. According to the actual situation, choose the product and structure of the appropriate model and specification, and appropriately set the connection mode between the components and the arrangement and installation position of the components. It is necessary to fully understand whether there are temperature changes and mechanical vibration problems in the area where the selected components are located. In order to ensure the safety and reliability of the electrical control system, it is first necessary to ensure that the devices in the system can communicate with each other.

(3) The main points of electrical control

In the electrical control system, the main key points are the PLC control circuit and logic gate group. Its role is to process and convert input/output information. The signal conversion process is completed by changing the input voltage or current; the value of the set parameter is sent to the PID algorithm controller. The specified value is compared with the actual situation to determine the deviation and generate a corresponding number of instructions to meet the electrical control. The requirements for the various functions of the system is to ensure that the system is stable, reliable, and running well during operation.

3 Construction of Experimental Platform for Three-Motor Synchronous Control System

3.1 Experimental Platform Hardware Structure

The experimental platform of the three-motor synchronous control system discussed in this article is mainly composed of two parts: the mechanical part and the transmission part of the synchronous control system. The electrical control part of the synchronous control system. The mechanical part is mainly the pillar of the synchronous meshing

gear, and the gear part is composed of various tension wheels and belts. The tension and transmission speed of the tension belt are the control objects of the experiment. The electrical control part completes the basic electrical control, gearbox and execution algorithm platform.

For this experimental platform, the control part of the synchronous control system should be configured with hardware or software respectively. After the configuration is completed, the control algorithm can be designed and written. The hardware configuration mainly includes the hardware configuration of Siemens S7-300, MPI communication between S7-300 and computer, and PROFIBUS communication between S7-300 and MM440 converter. The software configuration is mainly composed of STEP7 software configuration and WinCC configuration monitoring software.

3.2 Program Realization of Control Algorithm

In this experiment, the PLC part uses structured ladder diagram LAD programming, which mainly realizes motor drive and communication, sensor signal acquisition, data access and transmission. WinCC programming control and monitoring interface. The main part of the control algorithm is implemented by MATLAB/Simulink, and the data input and output use the OPC Write and OPC Read modules in the OPC tool. WinCC is used to set up the monitoring interface on the one hand and as an OPC server on the other.

3.3 Experimental Method

This article mainly carried out three kinds of electrical control system experiments based on adaptive fuzzy algorithm, including speed tracking, speed voltage decoupling experiment and voltage observation experiment. Check the speed synchronization of the three motors by setting the reference voltage value to zero. In theory, regardless of the amount of forward slip, the speed of the second motor is equal to the speed of the first motor, and the speed of the third motor is equal to the speed of the second motor. Check the synchronization of PID algorithm and adaptive fuzzy algorithm with square wave or sawtooth wave.

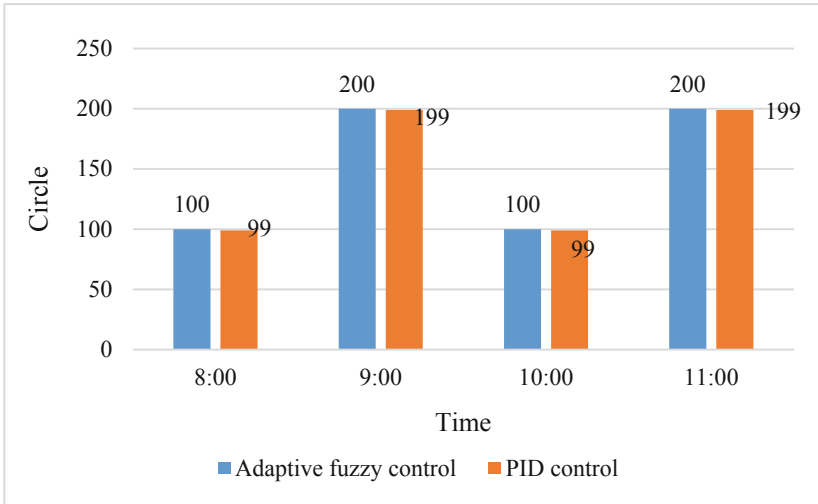
4 Analysis of Experimental Results

4.1 Three-Motor Speed Synchronous Tracking Results

Assuming that the tension value is 0, the speed results obtained by the adaptive fuzzy control algorithm and PID algorithm in the electrical control system are shown in Table 1. The data in the table represents the number of rotations of the motor per minute in different time periods.

Table 1. Three-motor speed synchronous tracking results

	Adaptive fuzzy control	PID control
8:00	100	99
9:00	200	199
10:00	100	99
11:00	200	199

**Fig. 2.** Three-motor speed synchronous tracking results

As shown in Fig. 2, we can see that in these two algorithms, the PID control value is always smaller than that of the adaptive fuzzy control algorithm, which shows that the adaptive fuzzy control has better capabilities. In the three motors, their tracking data is not displayed. It can be known that the adaptive fuzzy control algorithm has a better control effect.

5 Conclusion

With the continuous progress of my country's industrial production, higher requirements are put forward for electrical equipment and control systems. Therefore, we must improve electrical control technology. The electrical control of this design is based on the single-chip microcomputer as the core, through the analysis of its internal structure, and combined with modern advanced technology to carry out the research of the computer control system. Through the experiments in this paper, it is found that the electrical control system based on the adaptive fuzzy control algorithm can have the ideal control effect, which proves the feasibility of this method.

References

1. Sorial, R.R., Soliman, M.H., Hasanien, H.M., Talaat, H.E.A.: A vector controlled drive system for electrically power assisted steering using hall-effect sensors. *IEEE Access* **9**, 116485–116499 (2021). <https://doi.org/10.1109/ACCESS.2021.3105609>
2. Popkov, Y., Popkov, A., Dubnov, Y.: elements of randomized forecasting and its application to daily electrical load prediction in a regional power system. *Autom. Remote. Control.* **81**(7), 1286–1306 (2020). <https://doi.org/10.1134/S0005117920070103>
3. Das, L.N., Gupta, S.: Electrical power system transmission quality and power supplier micro grid control functional reliability. *Int. J. Syst. Assur. Eng. Manag.* **11**(2s), 325–328 (2020). <https://doi.org/10.1007/s13198-019-00943-1>
4. Muñoz-Vázquez, A.J., Sánchez-Torres, J.D., Parra-Vega, V., Sánchez-Orta, A., Martínez-Reyes, F.: A fractional super-twisting control of electrically driven mechanical systems. *Trans. Inst. Meas. Control* **42**(3) 485–492 (2020)
5. Gorostiza, F.S., Gonzalez-Longatt, F.M.: Deep reinforcement learning-based controller for SOC management of multi-electrical energy storage system. *IEEE Trans. Smart Grid* **11**(6), 5039–5050 (2020)
6. Litvinenko, A.M., Baranov, D.S.: An adaptive control system for a winding-machine electric drive. *Russ. Electr. Eng.* **91**(10), 620–625 (2020). <https://doi.org/10.3103/S1068371220100077>
7. Berdiev, U.: Improvement of efficiency of “O’zbekiston” series electric locomotive control system. *J. Tashkent Inst. Railway Eng.* **15**(4), 6 (2019)
8. Doyle, A., Muneer, T.: Energy consumption and modelling of the climate control system in the electric vehicle. *Energy Explor. Exploit.* **37**(1), 519–543 (2019)
9. Shi, K., Chen, Y., Zhou, H., et al.: Design of permanent magnet synchronous motor control system for electric vehicle air conditioning compressor based on vector control. *Open Access Library J.* **06**(1), 1–9 (2019)
10. Schmalfluss, L., Hahne, J., Farina, D., et al.: A hybrid auricular control system: direct, simultaneous, and proportional myoelectric control of two degrees of freedom in prosthetic hands. *J. Neural Eng.* **15**(5), 056028.1–056028.10 (2018)
11. Gryzlov, A.A., Grigor’ev, M.A.: A Control System of an electric drive with a synchronous reluctance machine with independent excitation. *Russ. Electr. Eng.* **91**(5), 307–311 (2020). <https://doi.org/10.3103/S1068371220050053>
12. Li, J., Zhou, Q., He, Y., et al.: Driver-identified supervisory control system of hybrid electric vehicles based on spectrum-guided fuzzy feature extraction. *IEEE Trans. Fuzzy Syst.* (99), 1 (2020)
13. Litsin, K.V., Tsukanov, A.V.: Automated electric drive for the control system of two-coordinate welding machine. *Izvestiya Vysshikh Uchebnykh Zavedenij. Chernaya Metallurgiya* **64**(5), 382–388 (2021)



Intelligent Optimization Algorithm in Virtual Design of Landscape VR

Danping Yi¹(✉) and Quannan Wang²

¹ Wuhan Institute of Design and Sciences, Wuhan 430000, China

² Hanseo University, Seosan, Korea

Abstract. In today's society, people have higher and higher requirements for the quality of life, especially for the comfort of the living environment. And VR virtual reality technology can well solve the needs of residential environment. When planning the entire scene in landscape VR design, the overall needs should be fully considered, and the needs of users should be met as much as possible. In order to improve the ability of landscape design, it is necessary to conduct research on intelligent algorithms. Therefore, this paper studies the application of intelligent optimization algorithm (IOA) in landscape VR virtual design through experimental methods and sampling tests. The data shows that after applying the IOA, the spatial layout accuracy rate of the virtual platform is up to 92%, and the response time is 0.8 s, which is the shortest time among the other three functions. This shows that the application of IOAs in landscape VR virtual design has great advantages.

Keywords: Intelligent optimization algorithm · Landscape virtualization · VR design · Application research

1 Introduction

VR technology has developed rapidly in recent years, and its application range is wide, and it can be applied in various fields. The goal of landscape VR design is to provide users with a real, accurate, efficient and real-time interactive three-dimensional scene experience. It can simulate the virtual reality system to achieve optimized configuration and improve space utilization.

There are many people studying the application of IOAs in landscape VR virtual design. For example, in view of the problem that environmental art design cannot make users feel the real and clear effect of environmental art design, Wang Peng studied the application of 3D virtual VR technology in environmental art design [1]. Chen Jingjing said that virtual reality technology provides a new design method and a new way of expression for design [2]. Shan Sheng briefly summarized the VR virtual reality technology [3]. Therefore, this article intends to use IOAs to conduct in-depth research on landscape VR virtual design, hoping to verify the role of IOA.

This article first studies some basic concepts and related algorithms of IOAs. Secondly, it analyzes the feasibility of virtual reality technology in landscape design. Then the parametric design and realization of the virtual plant landscape are described. Finally, the test results are obtained through experiments and relevant conclusions are drawn.

2 Application of Intelligent Optimization Algorithm in Landscape VR Virtual Design

2.1 Intelligent Optimization Algorithm

IOA is a kind of simulation of human thinking process, and through the comprehensive analysis of the human brain, so as to obtain the optimal solution. It can propose new schemes based on actual conditions and problems. First, according to the initial settings of the system, determine the optimal objective function and constraint conditions. Secondly, the best solution is obtained through calculation and analysis. Finally, iterative operation is performed under certain accuracy requirements. The most important thing in the IOA is the global search method and the local traversal strategy problem [4, 5].

On the basis of computer technology, network communication and control theory, through processing a large amount of data, the optimization objective function required by different schemes is obtained. IOA is a multi-variable nonlinear programming problem. It can determine the optimal solution or global optimal value range that can be achieved by each layer of neurons in each functional module according to the interdependence between the factors in each subsystem. In terms of computer processing power, it can linearize the input variable to make it an optimal solution. For non-linear problems, higher accuracy and stability are required. There are mainly two types in engineering applications. One method is called artificial neural network. Another method is used to describe the minimum function of the error (such as noise) produced when there are non-linear or uncertain factors in the artificial system, which brings the actual result close to the expected result [6, 7].

First, the objective function must be optimized. In the design process, we need to determine the optimal solution based on the initially set model structure. The second is to establish a set of mathematical programming problems. The constraints are solved by inputting actual objects into the computer system. Then use the heuristic algorithm to find the best solution as the optimal decision point. And calculate the corresponding output value and output position vector, and finally obtain the required objective function value to get the global optimization result, so as to realize the control of the whole process and [8, 9].

(1) PSO optimization algorithm

For a problem or function to be optimized, we give its feasible solution space, and imagine that a particle swarm composed of a certain number of particles with velocity and position variables moves in the feasible solution space. These particles continuously adjust their own motion speed by judging their own historical best position and the historical best position of the entire particle swarm, thereby realizing an intelligent search for the entire feasible solution space. Similar to the foraging behavior of the entire bird swarm, all particles finally is possible to move to the position of the optimal solution. The particle swarm in the particle swarm optimization algorithm does not mean a simple combination of a group of particles. It is actually a topological structure composed of a group of particles and a network of interaction relationships between particles [10, 11].

(2) Basic particle swarm algorithm process:

- 1) Initialize the velocity and position of all particles in the D-dimensional search space.
- 2) Substitute the position of each particle in the scoring function to obtain the fitness value of each particle, and initialize the historical best individual and historical best overall according to the fitness value [12, 13].
- 3) Enter the main loop:

$$\begin{cases} \vec{a}_i \leftarrow \vec{a}_i + \vec{w}_i \\ \vec{w}_i \leftarrow \vec{w}_i + \vec{P}(0, \zeta_1) \otimes (\vec{Q}_i - \vec{a}_i) + \vec{P}(0, \zeta_2) \otimes (\vec{Q}_h - \vec{a}_i) \end{cases} \quad (1)$$

- 4) Calculate the particle fitness value, and update the individual historical best value and the global historical best value.
- 5) If the conditions are met (usually meeting the requirements of the fitness value or reaching the maximum number of iterations), end the loop.

(3) PSO algorithm with weight

Suppose the size of the particle swarm is $M+1$, and the dimension of each particle is $D+1$. The specific evolution process of particles is:

$$w_{ik}^{s+1} = q w_{ik}^s + z_1 R_{1k}^s (G_{ik}^s - \tau_{ik}^s) + z_2 R_{2k}^s (G_{ik}^s - \tau_{ik}^s) \quad (2)$$

Among them, i represents the i -th particle, k represents the dimension of the particle, and s represents the number of iterations. In each iteration, the entire particle swarm is sorted according to the fitness value, and the worst half is replaced with the better half, while retaining the best historical value of the original particle memory.

2.2 The Feasibility of Virtual Reality Technology in Landscape Design

In theory, all types of landscape spaces can be presented in a virtual reality environment. From its birth to the present, virtual reality technology has become a brand-new expression mode in the development of these decades. The development goal of virtual reality technology is very clear and determined, that is, to design a better quality solution; to adopt a lower-cost design device; and to have a faster processing instrument. With the continuous optimization and improvement of virtual reality technology, this technology will bring a breakthrough change in landscape design, which will completely change the designer's design mode and make the design plan more scientific and reasonable. With the help of virtual reality technology, designers can reduce their time to actually design drawings, free them from tedious work, and give them more time to analyze the effects and practicability of the design. The designer fully immerses himself in the virtual reality space with the help of virtual reality technology, allowing the designer to observe the design object from any angle, so as to have a deeper understanding of the scale and proportional relationship of the design landscape space, and make the landscape design plan more suitable. In line with actual needs, the design effect is better and perfect.

(1) The practical value of virtual reality technology applied to landscape design

The characteristics and advantages of virtual reality technology have significant value in the field of landscape design and can promote the improvement and development of landscape design. The interactivity, thoroughness, richness and perception of virtual reality technology can optimize the designer's landscape design plan and plan performance. From the perspective of landscape design, when designers use virtual reality technology for preliminary design, they can increase the perception, experience and interactivity of their design, and create a virtual space for their design and development, allowing designers to observe the design in real time. The effectiveness and rationality of the design plan are optimized and improved. From the perspective of project performance, the public can view the design effect from multiple perspectives, and fully understand the ideas and opinions of landscape architects, and increase the public's sensory experience. At the same time, with the help of virtual reality technology, the construction situation of the landscape project can be simulated, which can allow the construction party to understand the operation of the project as a whole, arrange the construction time reasonably, monitor the whole construction process and realize the successful completion of the landscape project, which can effectively reduce Unnecessary expenses.

(2) Advantages of virtual reality technology in landscape design

The most significant advantage of virtual reality technology is that it can improve the communication efficiency between Party A and Party B. The existing landscape design methods generally use hand-drawn drafts, computer renderings and simple text descriptions, which can easily cause communication difficulties between the two parties. Resulting in constant modification of the design plan, increasing the design cost and the designer's workload. Therefore, in the general sense, the design expression technique can no longer meet the needs of the design work. Through virtual reality technology, users can experience design effects in a virtual environment, find out where the problem lies, improve work efficiency, and solve the scale problem between abstract thinking and actual space. Existing computer technology has been unable to meet user requirements. Through this technology, we can do our utmost to reduce the impact of unreasonable planning during construction and cause rework. Designers use virtual reality technology to "visit" the construction site to better understand the overall project.

(3) Application of virtual reality technology in various stages of landscape design

Participate in research landscape design projects. The virtual reality technology is used in the preliminary investigation and analysis of landscape design projects to analyze the feasibility of the project as a whole, and play an important role in reducing key errors during project operation. To fully grasp the requirements of landscape projects, it is necessary to fully consider the design function requirements, design forms, economic factors and time factors, clarify the difficulties that may occur in the design process and create design highlights. Combined with virtual reality technology, landscape designers use its characteristics to conduct systematic simulation and exploration, and have an overall understanding of the project overview.

Conceptual design landscape plan. Assist the project to complete the preliminary investigation, virtual reality technology as an auxiliary technology, assist in the completion of the project conceptual design. The application of virtual reality technology in conceptual design is similar to the way that designers think while drawing by hand. The designer creates a conceptual model in the virtual reality environment, which not only gets rid of the limitations of traditional two-dimensional drawings, but also enables more detailed analysis and evaluation of the project.

The performance of the landscape design plan. Virtual reality technology allows decision-makers and the public to better integrate into the scene through a new way of landscape expression, view the design landscape effect, expand the public's vision from multiple angles, experience the designer's design concept, and interact with the designer.

2.3 Parametric Design and Realization of Virtual Plant Landscape

This system is intended to visually design the type, quantity and three-dimensional layout of virtual scenes, virtual plant models through a mouse or keyboard, and finally generate an ideal virtual plant landscape. At the same time, in the design process, the virtual scene of the design can be roamed in real time to achieve a real and reliable design effect. The system is mainly composed of three modules. Model information module, parametric design module and X3D document generation module. The system development environment is. Operating system Microsoft Windows XP, JAVA platform is jdk1.5.7_21, JAVA development platform is EclipseSDK3.6.3, database uses SQL Server2015, database connection driver uses Microsoft SQL Server 2015 Driver for JDBC, X3D browser uses Xj3D4.0 ORC, X3D file The editor is X3D-Edit5.2. Its architecture is shown in Fig. 1:

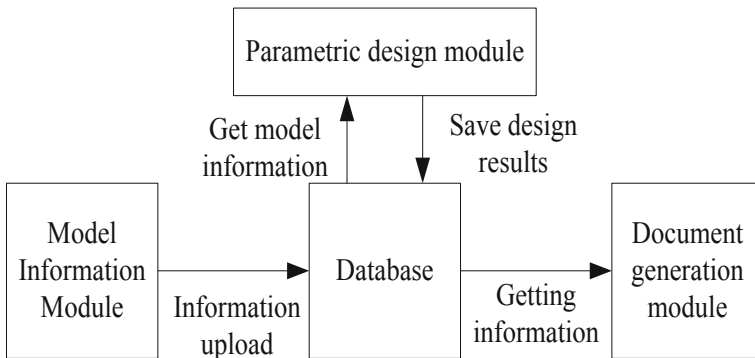


Fig. 1. Architecture of parametric design of virtual gun landscape system

The main function of the model information module is to upload and manage the relevant parameter information of the plant model and site model to the server and database, such as texture data, basic parameters, introduction information, etc. The server mainly saves the uploaded 3D model and various texture files, etc., The database mainly saves the access path of the 3D model and its texture data file on the server side and the basic parameter information of the model. Therefore, the design scheme of this module

is divided into three steps. The first step is to realize the site model, the plant model and its texture files and upload them to the designated folder of the server. The second step is to store the specific path of the model and its texture data file and the basic information related to it in the database on the basis of the smooth realization of the first step. The third step is model management, including the modification of introduction information, the adjustment of texture, and the setting of model LOD parameters.

3 System Development and Implementation

3.1 Data Collection and Processing

In order to objectively reflect the objective appearance of the natural and cultural landscape of Lushan Mountain, and integrate comprehensive tourism information data and special geoscience information, this paper has collected and processed a large number of relevant data. The process is divided into the following three aspects:

One is to obtain UAV panoramic point data and satellite remote sensing data through photogrammetry based on aerospace images. The second is data collection for natural and human landscape modeling. The third is to obtain fine texture data of picturesque buildings through surveying and mapping methods such as field surveys and close-up photography.

3.2 Data Database Construction

In order to facilitate the research and management of digital achievements, use the research and development of cultural heritage landscape resources and cultural tourism resources digital achievements database to organize, preprocess, manage and integrate multi-source data, and integrate multiple sources and basic multi-temporal geographic and heritage data. After the standardized and unified design of spatial data, a spatial data framework, networked data management and layered services, and historical data storage and management have been established.

3.3 Development of 3D Virtual Platform

The platform reproduces natural and cultural landscapes, and users can access the system by entering a website address in the browser. The main functions of the system include scenic spot overview, scenic spot introduction, roaming explanation, tourist routes, tourist facilities, real scene experience, geology and landforms, plant ecology, soil profile and scene control.

4 Analysis of System Test Results

4.1 Advantages of Intelligent Optimization Algorithms in the 3D Virtual Landscape Design Platform

In order to enhance the accuracy and efficiency of the development of the 3D virtual platform, this paper uses the proposed IOA. Introduce it into the platform for calculations,

and explore the accuracy and timeliness of the platform's scenic spots introduction, roaming explanation, real-world experience, and spatial layout. The results are shown in Table 1:

Table 1. Accuracy and timeliness data of the platform functions

	Accuracy	Timeliness
Introduction to scenic spots	89%	1.1
Roaming explanation	90%	1
Real experience	86%	1.5
Spatial distribution	92%	0.8

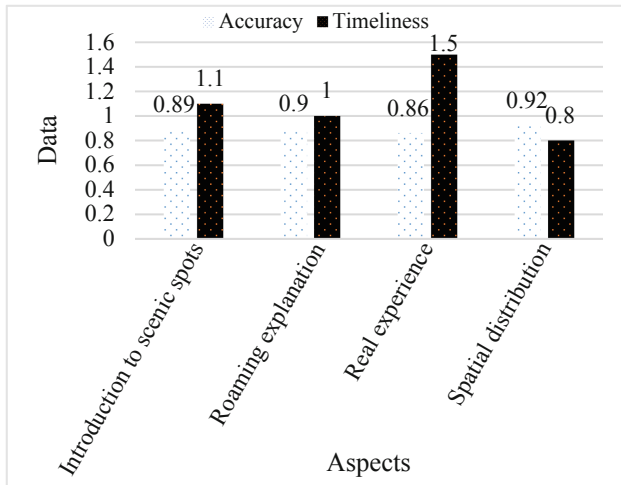


Fig. 2. Accuracy and timeliness data of the platform functions

As shown in Fig. 2, we can see that with the help of IOAs, the accuracy and timeliness of functions such as scenic spot introduction, roaming explanation, real scene experience and spatial layout are guaranteed. The accuracy level is maintained above 0.85, and the response time is about 1s.

5 Conclusions

In traditional design methods, we can know that when an object is disturbed by the outside world, its internal structure will change. In this case, it cannot be handled well. However, with the continuous development and progress of advanced science and technology such as computer technology and network communication technology, IOA continue to improve. Its application in various systems is becoming more and more frequent.

Using this algorithm can promote the functional improvement of the three-dimensional virtual landscape design system proposed in this paper. Experimental data shows that after the IOA is introduced into the system, the accuracy of the platform basically meets the requirements of landscape design. Therefore, the application of IOA has certain advantages for the system.

References

1. Peng, W.: Research on the application of 3D virtual VR technology in environmental art design. *Mod. Electron. Technol.* **041**(012), 168–171 (2018)
2. Chen, J.: Application of virtual reality technology in environmental art design. *Decor. World* **000**(023), 00249–00249 (2017)
3. Shan, S.: Design and implementation of VR virtual reality technology in virtual campus system. *J. Shandong Agric. Adm. Coll.* **037**(004), 30–32 (2020)
4. Nguyen, G.N., Viet, N., Joshi, G.P., et al.: Intelligent tunicate swarm-optimization-algorithm-based lightweight security mechanism in Internet of Health Things. *Comput. Mater. Continua* **66**(1), 551–562 (2020)
5. Pawara, R., Ahmad, I., Surana, S., Patel, H.: Computational identification of 2,4-disubstituted amino-pyrimidines as L858R/T790M-EGFR double mutant inhibitors using pharmacophore mapping, molecular docking, binding free energy calculation, DFT study and molecular dynamic simulation. *Silico Pharmacol.* **9**(1), 1–22 (2021). <https://doi.org/10.1007/s40203-021-00113-x>
6. Meng, C., Sun, M., et al.: Virtual design on detection system of sitting posture corrector. *Int. J. Plant Eng. Manag.* **24**(02), 29–38 (2019)
7. Moorthy, N.S.H.N., Brás, N.F., Ramos, M.J., et al.: Structure based virtual screening of natural product molecules as glycosidase inhibitors. *Silico Pharmacol.* **9**(1), 1–19 (2021)
8. Keene, D.: Beyond the VR hype-collective VR and the new technology landscape. *Sound Video Contract.* **36**(6), 56–58, 60 (2018)
9. Ali, J., Saeed, M., Tabassam, M.F., Iqbal, S.: Controlled showering optimization algorithm: an intelligent tool for decision making in global optimization. *Comput. Math. Organ. Theory* **25**(2), 132–164 (2019). <https://doi.org/10.1007/s10588-019-09293-6>
10. Ullah, A., Wang, B., Sheng, J., et al.: Optimization of software cost estimation model based on biogeography-based optimization algorithm. *Intell. Decis. Technol.* **14**(4), 1–8 (2020)
11. Ghathwan, K.I., Mohammed, A.J.: Intelligent bio-inspired whale optimization algorithm for color image based segmentation. *Pertanika J. Sci. Technol.* **28**(4), 1389–1411 (2020)
12. Sulaiman, M., Samiullah, I., Hamdi, A., et al.: An improved whale optimization algorithm for solving multi-objective design optimization problem of PFHE. *J. Intell. Fuzzy Syst.* **37**(3), 3815–3828 (2019)
13. Nayak, B., Choudhury, T.R., Misra, B., et al.: Estimation of component values of filter design using optimization algorithms. *J. Intell. Fuzzy Syst.* **37**(3), 3567–3579 (2019)



Processing Method and System of Artificial Intelligence Application Based on Open Source Deep Learning Framework

Yihan Tu^(✉)

School of Software and Blockchain, Jiangxi University of Applied Science,
Nanchang, Jiangxi 330000, China

Abstract. With the rapid development of computer technology, artificial intelligence has become a very important topic in the field of information technology in the world. In our living and working environment, machine learning is one of the methods with strong ability to process and understand information and high accuracy. It can transform complex data into simple and easy to store and manage algorithm tools that are characteristic of human brain memory and can quickly identify the behavior rules of target objects. It can also integrate and analyze a large number of original data to solve some practical problems. Therefore, based on the open source deep learning framework, this paper designs an artificial intelligence application processing system. Firstly, this paper introduces the concept and principle of deep learning, studies its application law, then expounds the application processing method of artificial intelligence, designs the framework of artificial intelligence application processing system, and tests its performance. Finally, the experimental results show that the artificial intelligence processing system based on deep learning algorithm has good performance in all aspects, the operation time and delay time are within 2–3 s, the occupied memory is about 3400k, and the CPU occupation rate is about 2%. This shows that the performance of the system has met the requirements.

Keywords: Open source framework · Deep learning · Artificial intelligence · Processing method

1 Introduction

With the popularity of the Internet, artificial intelligence is applied more and more, but because it plays an important role in real life, its development prospect is very broad. In recent years, various intelligent terminals have been produced based on the combination of network platform, biotechnology and big data mining [1, 2]. Based on neural cells and self-learning engine, a series of processing products with independent analysis ability, strong real-time performance and high reuse rate have been developed. These systems can provide users with more convenient, efficient and practical services that can meet people's personalized needs [3, 4].

Many scholars have conducted relevant research on deep learning. With the rapid development of computer network technology, machine learning has been widely studied and applied abroad based on the Internet platform. Among them, the United States, Japan and other developed countries have begun to develop artificial intelligence systems with independent intellectual property rights. The in-depth exploration of intelligent processing in China began in the late 1980s. Domestic scholars also actively invested a lot of resources to do relevant work. The Chinese Academy of Sciences developed a nonlinear mapping model neural network algorithm composed of a variety of sensors and capable of multivariable analysis, and applied it to in-depth learning, making a major breakthrough and progress in this regard [5, 6]. Some scholars have proposed that the “machine learning based on decision support” method can effectively solve the problems of insufficient ability and slow convergence speed of traditional artificial neural networks in processing high-dimensional data. In addition, some scholars use BP algorithm to predict and obtain an uncorrelated function relationship between the optimal solution and accuracy. It is found that when the training set is 1, the error is the smallest. With the continuous optimization and improvement of BP algorithm, better results can be achieved [7, 8]. The above research has laid the foundation for this paper.

Based on the open source deep learning framework and its related technologies, this paper designs an object-oriented processing process, and gives the corresponding classification method. In the system, the user pre scores according to the needs to obtain the original data set, and then uses the neural network algorithm to map the input to be identified into a standard function.

2 Discussion on the Processing Method of Artificial Intelligence Application Based on Open Source Deep Learning Framework

2.1 Deep Learning

2.1.1 Concept

The concept of deep learning is a network-based algorithm. In this process, complex data sets need to be decomposed into small sample sets and representative feature point structures. Deep learning is different from the relatively shallow learning concept of artificial neural network. Deep learning really realizes the idea of multi-layer neural network and emphasizes the depth of the network. The construction of multi-layer network enables complex function problems to be obtained by setting fewer parameters. Deep learning refers to processing data with machine language, The problem is solved through the results of analysis and calculation [9, 10]. Deep learning highlights the importance of feature learning, and does not change the original information of the data while changing the dimension, so as to make the learned data easier to process and apply. Clustering technology is commonly used in our life. It is based on artificial intelligence technology to process, analyze, summarize and draw conclusions on data and knowledge. Its main purpose is to discover the potential existing or possible changes or new features in the environment through the automatic identification of the computer. It gathers objects of different types, different types or similar degrees together to form a new individual (i.e. “person”), at the same time, it can also form a group (i.e. “group”) of people with great

differences in the same category or attribute, and finally judge whether there is similarity and correlation between targets according to the classification criteria [11, 12].

2.1.2 Features

The main feature of deep learning is that when training data, it can build a targeted, diverse and personalized knowledge base according to different types and characteristics. At the same time, there are many layers of neuron units in the structure of deep neural network. When there are many uncertain factors in complex data, which make the results unable to accurately predict the error probability or large deviation in the target mode, the multi-dimensional scale method can be used to solve it. In the process of learning, we can build a targeted, diverse and personalized knowledge base according to different types of characteristics. The process of in-depth learning is as follows:

(1) The bottom-up unsupervised learning is different from the traditional neural network. The feature learning process of the machine is an unsupervised learning process. In the training, the uncalibrated samples are used to train the bottom layer of the hidden layer, cooperate with some parameters of the previous layer to obtain better features than the input, and the previous output is used as the input of the next layer.

(2) The top-down supervised learning uses the sample data with artificial labels to train, uses the parameters of each layer obtained in the first step of unsupervised learning, and then adjusts the parameters of the whole model, and the error is transmitted from top to bottom

2.1.3 Algorithm

Logistic regression belongs to a typical discriminant model. The most intuitive performance is to ignore the joint distribution of data and model conditional probability, which is obviously different from the traditional generative model Bayesian network. The generative model first models the joint probability distribution of data, and then calculates the posterior probability of each data set through Bayes and other formulas. From a mathematical point of view, the logarithm of probability comes from the linear combination model of predictive variables, and the process of logical regression is divided into three steps. Firstly, the prediction function H is constructed. The essence of logistic regression is a classification algorithm, which is mainly used for two classification and some multi classification problems. Therefore, it uses the sigmoid regression function, and the expression of the function is:

$$G(z) = \frac{1}{1 + e^{-z}} \quad (1)$$

Thus, a prediction function $H(x)$ is constructed, which represents the probability that the result takes 1.

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \quad (2)$$

It can deal with nonlinear problems through the relationship between input data sets, implicit parameters and output results. It has good stability and robustness. At the same

time, it also reduces the computational complexity and improves the operation speed to a certain extent.

2.2 Artificial Intelligence Application Processing Method

The application method of artificial intelligence is based on machine learning. According to different objects, many different types of algorithms can be used to deal with these problems. For example, artificial neural network, fuzzy set theory and decision tree model. Among the current AI application processing methods, the method based on deep learning framework and SAR support system has good results. The application of artificial intelligence is mainly through computer and machine learning. In dealing with complex and fuzzy data, neural network and other methods are used for analysis. It has strong adaptive ability and fault tolerance, which is stronger than other computer technology fields. At the same time, it can quickly converge in the case of noise interference, and obtain the optimal solution through multi-layer processing of data. Therefore, it has a wide range of applications and obvious effect. It is a typical nonlinear mapping system. It establishes the model and gives the prediction results by dealing with the relationship between the sample points. When the training set reaches the required goal, it will produce the output value, which is different from the corresponding quantity in the real world. The application of this method is to combine multiple items to form a large file group to improve the classification efficiency and reduce the amount of calculation. The algorithm can effectively reduce the waste of resources and the increase of computing cost caused by redundant tasks in the process of machine learning.

2.3 Open Source Framework

Java open source framework is an application mode of algorithm based on deep learning. It uses object-oriented method, automatic classification technology and self-organization theory to gather complex data into the same class, and realizes the target set features with strong correlation between data, rich shared resources and multi-dimensional feature description ability. In computers, the method we usually use is the Java open source framework. It has three advantages: (1) it has strong scalability and can provide more services for users. (2) Because of its high reliability and security, it is widely used in various fields. For example, some data analysis software is applied in artificial intelligence processing system to help users make decisions, and this information is often called or modified text files and stored when we need it. (3) Open source framework is easy to manage and maintain, and can well solve the current problems. Hibernate is one of the excellent open source Java object relational mapping frameworks. It can not only replace most JDBC codes in Java applications, but also be integrated into Java EE system as a persistence layer framework. Hibernate can be well integrated with a variety of web servers or application servers. Now it supports almost all popular database servers. As a new ORM mapping tool, hibernate not only provides mapping from Java classes to data tables, but also provides data query and recovery mechanism. Compared with using JDBC and SQL to manually operate the database, using hibernate can greatly reduce the workload of operating the database.

3 Experiment

3.1 Framework of Artificial Intelligence Processing System

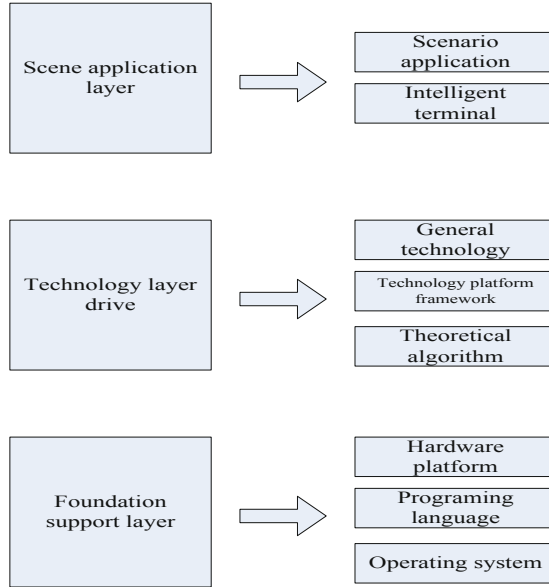


Fig. 1. The AI processing system framework

The application framework based on the depth of open source is mainly composed of three parts, as shown in Fig. 1: scene application data collection, technology driven system management and basic analysis and processing. In the process of high-capacity operation, a variety of sensors are used to obtain the original source code information. Through the extraction of different characteristic signals, the multi-channel output signals are obtained. The source code information is obtained by using web crawler technology and uploaded to the local database to realize the operation modes such as real-time packet capture and classification, and complete the task target identification, so as to ensure the scalability of system management function, operation platform function and maintenance and upgrading.

3.2 Experimental Data Collection

In this paper, artificial neural network is used as the main method of data mining. Different models are used to train classifiers and process them according to samples. Firstly, we preprocess the data set and select an appropriate number, type and characteristic parameters (such as noise intensity) in the input set. During data collection, we need to preprocess the collected content, parameters and other information, input all the data saved in the original database, and then download relevant files from the Classification

Library through automatic retrieval technology, And return to the user, extract and store the corresponding type of label documents according to the text file, and then output them to the clustering algorithm. Finally, use this result to provide recommendation services for the user.

3.3 Experimental Test Process

In this application experiment, we use five different parameters to verify. It includes training set generation, test sample preprocessing and experimental data analysis. Because this paper is used to support the multi-user online system of deep learning algorithm, it is necessary to segment the data appropriately, and other parameters may affect the final result. In the training process, first map the data set to be tested to a set of eigenvalue functions, and then select some specific types of samples from these data sets as the final goal. Because each classifier corresponds to different category attributes, we can group the selected samples to realize hierarchical processing method and crossover operation. Finally, the interaction design and after the interaction, the activation function value is returned, which is realized and tested by simulation software.

4 Discussion

4.1 System Performance Test and Analysis

Table 1 shows the performance test data of artificial intelligence processing system.

Table 1. Performance data testing

Test times	Performance period (s)	Delay time (s)	CPU occupancy rate (%)	Memory consumption (k)
1	3	1	2	3142
2	4	2	3	3254
3	2	1	2	3230
4	3	1	1	3415
5	2	1	2	3243

When processing data, the method based on open source deep learning framework mainly obtains the optimal solution by analyzing a large number of training sets. However, due to too much manual intervention, inconvenient control and multi-threaded operation, complex process is used. In this paper, iterative strategy is used to achieve fast convergence and local search ability. As can be seen from Fig. 2, the artificial intelligence processing system based on deep learning algorithm has good performance in all aspects during operation, the operation time and delay time are within 2–3 s, the occupied memory is about 3400k, and the CPU occupation rate is about 2%. This shows that the performance of the system has met the requirements.

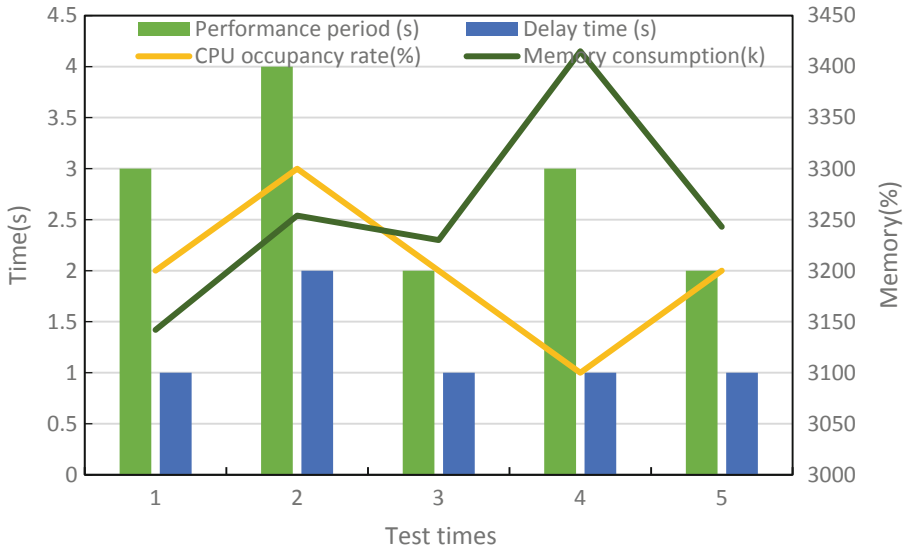


Fig. 2. Performance testing

5 Conclusion

With the development of the Internet, artificial intelligence will be widely used in the future, which also solves many practical problems for us, such as high learning efficiency, strong autonomous control ability and so on. This paper introduces a model of data processing, text classification and decision analysis based on the combination of open source deep learning framework and machine learning methods, analyzes and expounds the processing process based on deep model from different angles, solves the practical problems of artificial intelligence processing system, and gives its application prospect in the future.

References

1. Wang, Y., Yang, F., Zhang, J., Wang, H., Yue, X., Liu, S.: Application of artificial intelligence based on deep learning in breast cancer screening and imaging diagnosis. *Neural Comput. Appl.* **33**(15), 9637–9647 (2021). <https://doi.org/10.1007/s00521-021-05728-x>
2. Cheng, J., Song, T., Liu, Z., et al.: Artificial intelligence technology based on deep learning in digestive endoscopy imaging diagnosis. *Pers. Ubiquit. Comput.* 1–6 (2021)
3. Miyake, J., Ohigashi, H., Niikoka, H., et al.: Medical application of artificial intelligence/deep learning. *Brain and nerve = Shinkei kenkyū no shinpo* **71**(1), 5–14 (2019)
4. Anxie, T., Bing, L.: Application of deep learning and artificial intelligence in the psychological mechanism of language activity. *J. Intell. Fuzzy Syst.* **38**(23), 1–13 (2020)
5. Shi, M.: Knowledge graph question and answer system for mechanical intelligent manufacturing based on deep learning. *Math. Probl. Eng.* **2021**(2), 1–8 (2021)
6. Tsai, M.-F., Huang, J.-Y.: Sentiment analysis of pets using deep learning technologies in artificial intelligence of things system. *Soft. Comput.* **25**(21), 13741–13752 (2021). <https://doi.org/10.1007/s00500-021-06038-z>

7. Yopez, J., Castro-Zunti, R.D., Ko, S.B.: Deep learning-based embedded license plate localisation system. *Intell. Transp. Syst. IET* **13**(10), 1569–1578 (2019)
8. Ha, A., Aa, A., As, A., et al.: The promise of artificial intelligence and deep learning in PET and SPECT imaging - ScienceDirect. *Physica Med.* **83**, 122–137 (2021)
9. Zhang, W., Li, H., Li, Y., Liu, H., Chen, Y., Ding, X.: Application of deep learning algorithms in geotechnical engineering: a short critical review. *Artif. Intell. Rev.* **54**(8), 5633–5673 (2021). <https://doi.org/10.1007/s10462-021-09967-1>
10. Lee, Z. Y., Karim, M.E., Ngui, K.: Deep learning artificial intelligence and the law of causation: application, challenges and solutions. *Inf. Commun. Technol. Law*, 1–28 (2021)
11. Salvaris, M., Dean, D., Tok, W.H.: Deep Learning with Azure (Building and Deploying Artificial Intelligence Solutions on the Microsoft AI Platform) || *Trends in Deep Learning*, pp. 53–75 (2018). <https://doi.org/10.1007/978-1-4842-3679-6>
12. Hori, K., Ikematsu, H., Shinmura, K., et al.: The diagnostic artificial intelligence (AI) system for detection of colon polyps with high efficiency of deep learning. *Gastrointestinal Endoscopy* **87**(6), AB240–AB241 (2018)



Application of Computer Virtual Technology in Physical Training Video

Taowen Wei^(✉)

Anhui Vocational College of Metallurgy and Technology, Ma'anshan, China

Abstract. The physical training of virtual computer technology is now a new hot spot in the field of international sports science, and it has received more and more attention from domestic sports research institutions. This article studies the application of computer virtual technology in physical training videos, and understands the related theories of physical training and computer virtual technology on the basis of literature data, and then designs a physical training system based on computer virtual technology. Explains the application of computer virtual technology in physical training, and then tests the designed system. According to the test results, the average frame rate of the system designed in this paper has reached 66 or more, indicating that it meets the basic requirements of the system.

Keywords: Virtual technology · Physical training · Computer technology · Sports science

1 Introduction

Monitoring the training process of athletes has become a part of modern physical training [1, 2]. With the development of modern sensors and computer technology and people's further understanding of the scientificization of sports training, intelligent monitoring in the training process of athletes has also been paid more and more attention [3, 4]. Nowadays, some advanced sports teams in the world have taken the lead in making physical training intelligent, and some of them have successfully realized it [5, 6]. Virtual training is now a hot topic in the field of sports, and many foreign researchers in the field of sports are exploring this aspect [7, 8].

Based on long-term research on physical fitness training, relevant scholars have found that compared with developed countries with foreign physical fitness training, China's physical training is still in the initial stage of traditional training, sports training concepts and theories are relatively backward, and the method innovation is insufficient. There is no systematic training for quality coaches, and even lack of coaches [9]. However, some scholars believe that modern sports technology is a systematic engineering developed by the integration of multiple disciplines. It is a synthesis of neurophysiology, functional anatomy, biomechanics, and sports medicine. It is also basic research based on evolution, human development, and sports models. Therefore, physical training has become a major comprehensive curriculum in the fields of modern sports teaching and health education in China [10]. Some researchers believe that due to the rapid development of

big data analysis, artificial intelligence and computer technology, modern sports technology has entered a new era of informatization and digitization. Significant changes have taken place in traditional training concepts, training methods, training systems, and competitive mode training. “No data, no training” and “No monitoring, no training” are the basic principles of high-level competition. The training team monitors, transmits and optimizes the entire process of technical and tactical, physical fitness, status, stage, and management, and transforms traditional and empirical training into information, digital, and scientific training. Data science is becoming the supporting theory for winning championships on the field [11]. In summary, the digital trend of physical training is already obvious. With the development of technology, virtual technology will also be a helper for physical training.

This paper studies the application of computer virtual technology in physical training, analyzes the digital trend of physical training and the advantages of computer virtual technology on the basis of literature data, and then designs a physical training system based on computer virtual technology. To illustrate the application of computer virtual technology in physical training, and finally tests the designed system, and draws relevant conclusions through the test results.

2 Research on Physical Training and Computer Virtual Technology

2.1 The Digital Trend of Physical Training

The numerical demonstration of sports training is nothing new. Because part of the field of sports training research is to use numerical values to measure the training process and summarize it through the law. In competitive sports, the collection and classification of human movement data is the main core of the entire sports analysis [12]. Athletes' physiological and biochemical processes also require the measurement of the athlete's internal load in order to achieve effective management of the athlete's functional status and fatigue, and the sports team training methods and sports biomechanics process require the measurement of the athlete's external performance during the exercise and control the athlete performance management of athletes. Due to the constraints of training equipment and technical means, it is difficult to monitor in time during the training process and adjust the training content of the players in time through feedback information. For example, middle- and long-distance runners can accurately track the athlete's workload through the heart rate monitor, and cyclists can also monitor the athlete's work through the heart rate monitor or power meter. However, physical training includes various forms of sports and complex skills, making digital detection more troublesome.

2.2 Advantages of Computer Virtualization Technology

- (1) Immersive experience: It is the key point of VR technology experience. This means that users can feel the fidelity of the virtual environment, especially in the process of vision, hearing, touch, smell, taste, and body immersion. Therefore, simulation training usually uses a three-D battlefield environment image resource library, including historical background, environment, characters, and a sound source library that can simulate different sound effects, and uses real-world simulation to improve the quality of training.

- (2) Interactivity: refers to the natural degree of people with experience in the virtual reality system to deal with things, and the natural degree of acceptance of human emotions and reactions in the virtual reality system, etc. The creation of this type of interaction mostly uses data bracelets, head-mounted displays, 3D mice, and various specialized 3D interactive devices. Therefore, in the simulation training, participants use multiple virtual weapons and equipment. These devices can evaluate and respond appropriately in time, showing the characteristics of direct interaction.
- (3) Imagination: VR technology aims to immerse people in this environment and acquire new knowledge through the creation and construction of virtual environments. This will generate new ideas and unleash your imagination. For example, simulation training can improve the perceptual and rational understanding of the training environment and improve the skills in the real world through the experience and operation of virtual and virtual equipment.

2.3 Virtual Technology Algorithm

The ROAM algorithm does not store a huge array of triangle coordinates to describe the terrain, but mainly uses a binary triangle tree to maintain the coordinates of the triangle, and uses a tree-based structure to maintain the depth of the triangle. The amount of control memory increases exponentially as it grows. In the binary triangle tree, each node is an isosceles right-angled triangle with no affiliated nodes. Separating from the triangle vertex at the center of the hypotenuse creates two new right-angled isosceles triangles, and then the two right-angled isosceles triangles form auxiliary nodes of the original node. This subdivision process allows to go back and repeat the required level of detail. ROAM can divide small squares very quickly and update them dynamically.

Modeling requires error calculation. The calculation method is as follows.

$$SplitMetric = (Variance * MAPSIZE * 2) / Distance \quad (1)$$

In the expression, Variance is the deviation of the node, and MAPSIZE is the width of the entire ground grid. This resolution error is relatively easy to calculate, taking into account the surface roughness and the distance between the observation point and the node. If the result of splitMetric settlement is greater than the global variable G defined by the system, the system will continue to separate the nodes, otherwise, the split will stop. This algorithm guarantees a constant frame rate by determining the value of G.

The strategy of this paper to improve the ROAM algorithm is to dynamically divide large plots of land, and divide the ground into smaller squares before dividing the large plots of land, then use the quadtree algorithm to divide the quadtree blocks. For the four-element tree algorithm, it provides a special definition and division of the four-element tree. The algorithm uses a relatively simple and fast calculation method to divide it into many squares of different sizes. The size of the square block depends on the distance from the viewpoint. The further away from the viewpoint, the larger the sub-parts obtained, and the closer to the viewpoint, the smaller the sub-parts obtained. Set whether the quadrilateral node is divided, where D is the side length of the topographic map, W is the side length of the small square, and L is the distance between the viewing angle and

the node requirement is

$$W \geq kD/L \tag{2}$$

Where k is a constant.

3 Physical Training System Based on Computer Virtualization Technology

3.1 Establishment of Virtual Human Model

Since the human body is a very complex and easily deformable body, not only the static shape of the human body, but also the skin deformation of the human body shape during the human body movement must be considered in the study of the human body shape. So far, the most commonly used human modeling method in virtual environments is hierarchical modeling technology. In this layered model, the virtual human body model is composed of a basic bone layer, a muscle layer, a skin layer, and some specific environment layers, which represent decorative elements of the human body, such as virtual facial hair and clothes. The basic skeleton is determined by the motion data of the human skeleton, which determines the basic posture of the human body. The muscle layer determines the deformation of various parts of the human body, and the skin deformation is affected by the muscle layer. Finally, the skin layer determines the appearance of the virtual face. Figure 1 shows a modeling method, which is a practical personalization. The method is divided into three steps: driving model, matching model and merging model.

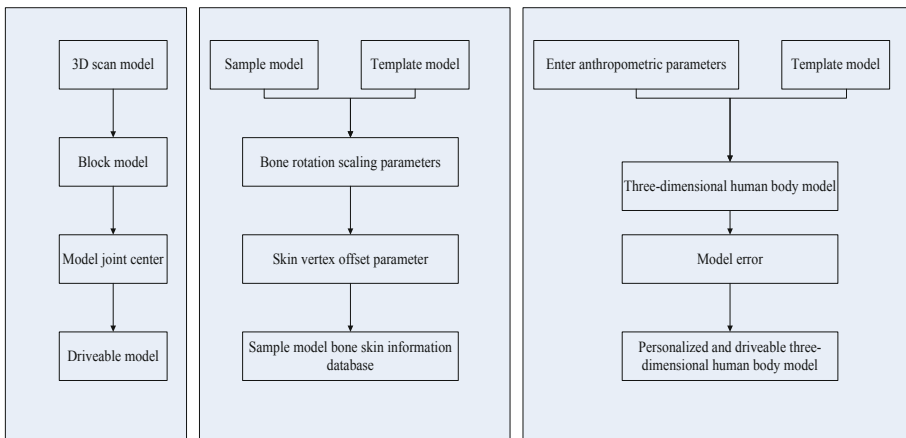


Fig. 1. Modeling method flowchart

The model-based stage first follows the Morse principle, constructs a Morse function based on the earth distance, calculates the equidistant area of the earth distance, and extracts the topological structure from the scanning model at different locations. Next, analyze the roundness of the ground from the isosceles surface, and extract the joint center of the model. Finally, connect the top of the model to the bones to drive the scan model. This process has nothing to do with the position of the scanned model, and is a more common bone extraction and auto-induction method.

In the model matching stage, the model model is first exported, and the target model is matched with bones and skin. According to the rotation and scaling information of the bone chain, the skin peak of the model is deformed to achieve the bone matching of the model. After the bones are matched, the top of the skin of the model is projected onto the surface of the target model. The projection strategy is as follows: the projection point is closest to the starting point, and the normal direction of the triangle where the projection point is located is close to the starting point. Finally, use the mass spring model to repair any errors that may have occurred during the process. After many repetitions, the skin matches the target model in the model.

In the model fusion stage, first in the measurement parameter space, according to the anthropometric input parameters and the measurement parameters of the sample model, use the radial-based function grid to calculate the fusion weight of each sample, and then merge the bone detachment information based on these weights. For skin deformation, according to the bone-based skin peptide deformation algorithm, the bones of the model model are scaled and the affected points of the bones are deformed according to the bone scaling ratio to obtain the deformed model, and then all the models are compared and the skin displacement information of the model is combined, and create a new model skin transformation.

3.2 Virtual Scene Design

After completing the modeling of the virtual human body and the design and debugging of the controller, the design integration of the virtual reality world began. At this stage, the design of virtual scenes and physical exercise actions are directly related to the meaning and effect of users using virtual reality to complete physical actions. Therefore, the scene designed in this work is obstacle training, which is modeled based on conventional obstacle scenes.

3.3 Vision Sensor

In this article, Kinect will be used for visual capture. With its advanced bone monitoring function, Kinect can actively monitor one or two people and receive the 3D coordinates of up to 20 joint spaces of the human body, but the limbs must be extracted and processed to recognize the user's limb movement and may react to the next steps exercise. Kinect processes the data from the sensors in a "pipeline" manner and the most primitive information needed to obtain the complete skeleton data of the active tracking object is the detailed data stream. Therefore, the following steps are required from opening the device to retrieving bone data.

(1) Specify general response events

There are two ways to get the latest buffered bone data frame. The first method is to use the voting function to get the buffer inside the loop, and read the bone data immediately after it is ready. The second method uses the event function to set the bone-readable event and passes the Windows WaitForSingleObject() kernel object wait function. When new bone data is generated, the event is activated and the standby process ends. Currently call API related methods to get data and reset event ID.

(2) Start NUIAPI

When calling NuiInitialize() to initialize and start the Kinect sensor, you must also specify the corresponding flag for the required sensor data. Only after a specific data mark type is specified, the relevant sensor information can be used in subsequent procedures. There are five types of switcher logos: audio data logo, color data logo, depth data logo, depth data logo with user ID, and bone data logo.

(3) Open the sensor information channel

After initializing and specifying the sensor information type flag, open the corresponding feed information channel cache and read data. The basis of bone data generation is to separate the user from the depth map. Therefore, the user ID needs to be used here to open the depth data stream channel to generate more bone data information.

(4) Multithreading to read data frame

Receive data in the above event mode. When the detailed reading event is activated, the detailed reading and processing data alarm will be triggered, and the corresponding method for obtaining detailed data will be called. Skeleton data when detailed information is received, regardless of whether there is a user.

3.4 Development of Virtual Platform

A good virtual reality platform must have powerful functions, huge economy, and rich interface capabilities with external devices. Therefore, for this task, VC++ and OpenInventor 7.0 will be selected as the development platform for the virtual training system integrated with artificial intelligence.

Open Inventor 7.0 is a powerful C++-based 3D graphics production system. Interactive Visual Framework IVF class library can easily call all Microsoft MFC package class libraries and C/C++ programming environment, combined with powerful and rich 3D graphics editing API functions, effectively complete the input and output of external images. IVF is essentially a Microsoft MFC application that integrates Win32 API and includes SoWin API and MFC-based 3D graphics extension class library.

4 System Test

In order to detect the frame rate memory of the virtual scene of the system, the experiment selects the ground as a 1024 * 1024 virtual scene, and divides the ground into 64 blocks. The experiment obtains a series of different analysis effects by changing the error limit, and introduces traditional algorithms for comparison. The results are shown in Table 1:

Table 1. System test results

Error threshold G	Traditional algorithm	Improve algorithm
0.1	5.2	6.2
0.2	9.5	10.5
0.5	21.2	25.2
1.0	33.4	39.6
2.0	44.7	49.9
5.0	58.4	62.6
10.0	59	67

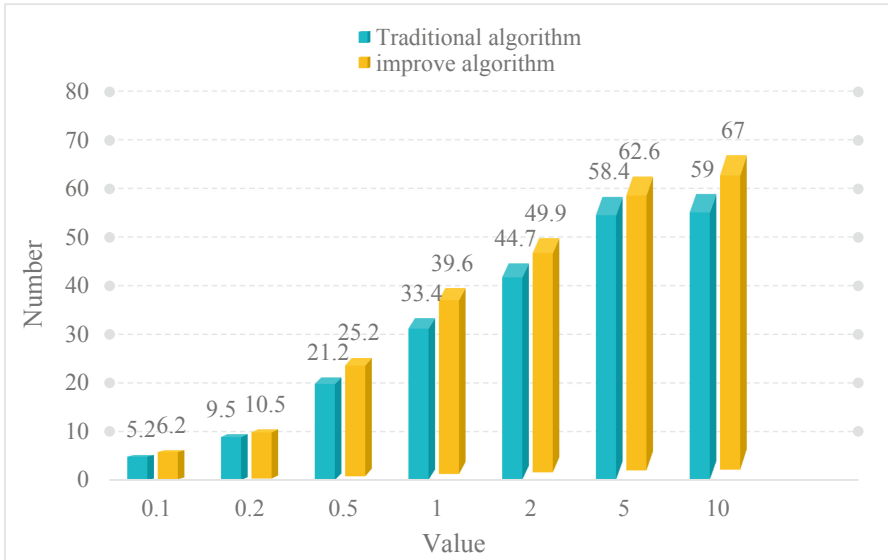


Fig. 2. System test results

It can be seen from Fig. 2 that after the upgrade, the frame rate has increased at the same resolution, but the improved algorithm still has its shortcomings. In the process of improvement, the soil roughness was not fully considered, and the influence of some simulated soils may not be obvious.

5 Conclusions

This article focuses on the application of computer virtual technology in physical training. After understanding the relevant theories, a physical training system based on computer virtual technology is used to explain the application of virtual technology in training, and then the designed system is tested and passed the test. The result is that the average frame rate of the system designed in this paper can reach 66 or more, which can meet the needs of the system.

References

1. Agnieszka, K., Monika, P., Sylwia, Z., et al.: Potential applications of virtual reality devices in older people. Narrative review. *J. Educ. Health Sport* **9**(6), 177–186 (2019)
2. Bertram, S., Brixius, K., Brinkmann, C.: Exercise for the diabetic brain: how physical training may help prevent dementia and Alzheimer's disease in T2DM patients. *Endocrine* **53**(2), 350–363 (2016)
3. Gombos, G.C., Bajsz, V., Pék, E., et al.: Direct effects of physical training on markers of bone metabolism and serum sclerostin concentrations in older adults with low bone mass. *BMC Musculoskelet. Disord.* **17**(1), 1–8 (2016)
4. Orr, R.M., Pope, R.P., Knapik, J.J.: A physical training framework for reserve personnel: a rationalization and recommendations. *Strength Cond. J.* **38**(4), 36–41 (2016)
5. Luger, E., Dorner, T.E., Haider, S., et al.: Effects of a home-based and volunteer-administered physical training, nutritional, and social support program on malnutrition and frailty in older persons: a randomized controlled trial. *J. Am. Med. Dir. Assoc.* **17**(7), 671.e9-671.e16 (2016)
6. Lewicka, A., Lewicki, S., Kłos, A., et al.: Influence of protein deficient diet, vitamin B2 supplementation and physical training on serum composition of polyunsaturated fatty acids (PUFAs) in rats. *Ann. Agric. Environ. Med.* **24**(2), 185–189 (2017)
7. Grier, T.L., Canham-Chervak, M., Anderson, M.K., et al.: Effects of physical training and fitness on running injuries in physically active young men. *J. Strength Cond. Res.* **31**(1), 207–216 (2017)
8. Campos, L.C.B., Campos, F.A.D., Bezerra, T., et al.: Effects of 12 weeks of physical training on body composition and physical fitness in military recruits. *Int. J. Exerc. Sci.* **10**(4), 560–567 (2017)
9. Shibata, S., Kawano, H., Maemura, K.: Effects of long-term physical training on the bearers of a float during the Nagasaki Kunchi festival. *Intern. Med.* **56**(1), 11–16 (2017)
10. Grijota, F.J., Muoz, D., Bartolomé, I., et al.: Arsenic, cadmium and lead erythrocyte concentrations in men with a high, moderate and low level of physical training. *Biol. Trace Elem. Res.* **195**(1), 39–45 (2020)
11. Yee, K.E., Moore, J.B., Grieve, G., et al.: Assessing value of physical training for tactical athletes: 1440 board #34 May 28 9:30 AM - 11:00 AM. *Med. Sci. Sports Exerc.* **52**(7S), 375 (2020)
12. Kapusta, J., et al.: The influence of hydrotherapy and physical training procedures on selected haemodynamic parameters of the circulatory system in patients with atherosclerotic peripheral vascular disease. *Polski merkuriusz lekarski: organ Polskiego Towarzystwa Lekarskiego* **46**(275), 213–216 (2019)



Machine Automatic Translation Quality Evaluation Model Based on Recurrent Neural Network Algorithm

Beibei Ren^(✉)

School of Translation Studies, Xi'an Fanyi University, Xi'an, Shaanxi, China

Abstract. In real life, the evaluation of translation quality is widely used. Not only translation activities need evaluation and comparison, but also translation quality evaluation in the editing and publishing of foreign language works, text translation teaching and other fields. The purpose of this paper is to study the quality evaluation model of machine automatic translation based on the recurrent neural network algorithm. First introduced the research background of this article, discussed the importance of research, and then briefly reviewed the research and development of machine translation evaluation, focusing on the current situation of recurrent neural network and evaluation method automation, and summarized the translation evaluation of automated machine technology. Finally, the main content of this article is summarized. The language model based on iterative neural network algorithm is studied. Basic experiments are carried out on the repetitive neural network translation quality evaluation model for machine translation quality evaluation. On the basis of basic experiments, 100 sentences were randomly selected to evaluate the performance of the translation system. The experimental results show that the translation quality estimation system of the machine automatic translation quality evaluation model based on the recurrent neural network algorithm established in this paper is 0.1 points better than the QuEst system.

Keywords: Recurrent neural network · Machine translation · Automatic translation · Quality assessment

1 Introduction

For the communication and dissemination of civilizations between different languages, language translation has become a key factor in the economic, political and cultural exchanges of various countries [1]. The so-called translation refers to a language activity. It includes both oral translation and written text translation. Essentially, it refers to a language that converts one language text into another without changing its meaning. The language behavior of texts in various languages. Strictly speaking, translation is a kind of “cross-cultural activity”. With the continuous development and progress of today’s society, facing a large number of various translations, this inefficient process has gradually become people’s daily translation activities, translation education and translation research. One obstacle, manual evaluation will become more and more inaccessible

[2, 3]. Therefore, evaluating the translation quality of human translation has become an indispensable feedback link in our daily translation process [4, 5].

The evaluation of machine translation is of great significance to machine translation research. Evaluation is based on the judgment of the evaluator, and it is necessary to ensure the reliability of the judgment [6]. Since the Z number has the potential to overcome the problem of information reliability, the introduction of this concept brings an improvement in describing uncertain information in multi-criteria decision making. Inspired by this development, Qiu D used Z-number to develop a comprehensive translation evaluation method. The specific steps of this method are given, and two different machine translation target texts from the same original text are studied using this method. The results show that the proposed method provides useful guidance for machine translation quality evaluation [7]. Fomicheva M designed an unsupervised QE method that does not require any training or access to other resources except the MT system itself. Unlike most of the current work that regards the MT system as a black box, they explore useful information that can be extracted from the MT system as a by-product of translation. Through the use of uncertainty quantification methods, they have achieved a very good correlation with human judgments of quality, which is comparable to the most advanced supervised QE model. To evaluate their method, they collected the first dataset that supports the work of QE's black box and glass box methods [8]. It is of practical significance to study the machine automatic translation quality evaluation model based on the recurrent neural network algorithm.

The research content of this paper is based on the recurrent neural network algorithm for machine automatic translation translation quality estimation method. By analyzing the related work of neural translation quality estimation, we understand the current status of quality estimation research using only the translation fidelity features of the translation, and propose a machine automatic translation translation quality estimation method based on the recurrent neural network algorithm to improve sentence-level quality Estimate the effect of model performance. Finally, experiments are carried out on tasks related to the comparative analysis of scoring results and language model performance analysis. The main research process is: preprocessing of parallel corpus and quality estimation evaluation corpus; implementing quality estimator based on BiLSTM network, and fusing context word vector and quality vector to build sentence-level neural translation quality estimation model; model training and optimization; verification model.

2 Research on Machine Automatic Translation Quality Evaluation Model Based on Recurrent Neural Network Algorithm

2.1 Recurrent Neural Network

Compared with the traditional neural network model, the biggest feature of the recurrent neural network is to spread the information contained in the node layer by layer through the iterative method of the hidden layer, so that the final top-level output node can learn rich feature information [9, 10]. In the sentiment analysis task, with the help of the recursive characteristics of the syntactic tree, on the input layer, we represent the words

in the sentence in the distributed representation of the word vector (here, we use the word2vec word vector representation method provided by Google) to The two-by-two merging of text syntax tree nodes continuously calculates the parent node vector layer by layer [11, 12]. Function f is the most basic activation function in neural networks, and its activation form uses function $f = \tanh$. In order to continue the recursive computing characteristics, the dimension of the hidden layer node in the recurrent neural network must be the same as the dimension of the initial leaf node, that is, the dimension of the word vector. Therefore, the dimension of the matrix W that stores the weight information needs to be set to. On each neuron node of the syntactic tree, we judge the distribution of emotion by applying a softmax function layer to the feature. The softmax() function is used to soften the output c -dimensional emotion level vector so that each component in the vector satisfies Probabilistic attributes, through the use of recurrent neural networks, we can see that the emotional information in the words (hidden in the features represented by the word vector, and displayed through the softmax layer), with the help of the syntactic tree structure, spread layer by layer, and communicate with each other. Fusion strengthens or rejects the reversal, and finally the emotional information is transmitted to the top of the syntax tree, completing the judgment of the degree of emotional tendency in the entire text sentence.

2.2 Model Evaluation Criteria

The simplest and most direct criterion for evaluating a language model is to apply the model to subsequent tasks, such as speech recognition, text classification, and machine translation. Although this method is intuitive and practical, it is necessary to construct a corresponding system for subsequent tasks, which greatly increases the experimental cost, and it is difficult to compare the performance differences of language models. Therefore, the current mainstream method of evaluating language models is to compare the parameters of the model itself. The two main methods are described below.

(1) Number of parameters

For any model, the space required to store itself and the computing resources required to train the model are determined according to the number of parameters, and the training time is positively correlated with the number of parameters. Sometimes the number of parameters will be very large, so that the model can better solve the problem. However, when training the model, a large amount of computing resources are correspondingly required, which is difficult to satisfy in actual production. Therefore, it is necessary to find a model with better performance under the same computing resources. For the N-gram language model, the number of N-gram sequences obtained on the training data is the number of parameters of the model, and the size of the training data set has a positive correlation with the number of parameters. In the neural network language model, the number of training weights in the network represents the number of parameters. Therefore, the number of parameters is only suitable for the performance comparison of models of the same type and different scales.

(2) Confusion

In information theory, perplexity is an index used to measure the pros and cons of predicted results, as well as probability models or probability distributions. Perplexity in the field of statistical language models is often used to evaluate models. The statistical

language model calculates the probability similarity between the given word sequence and the actual possible word sequence. If the model learns the real possible words, the better the performance of the model. At this time, the degree of uncertainty calculated by the model, that is, the degree of confusion, will be lower. Although the perplexity does not directly show the performance of the model, when comparing the performance differences between different types of language models, the perplexity can be used as a criterion.

3 Investigation and Research on the Quality Evaluation Model of Automatic Machine Translation Based on Recurrent Neural Network Algorithm

3.1 Operating Environment

The experiment in this paper is to accelerate the training speed of neural network language model, using a nVidiaGTX1060Ti graphics card as the hardware acceleration card. The operating system is Windows, which has good support and stability for all kinds of open source software used in the experiment. This chapter mainly consists of two parts: one is to train two language models, and then to examine the performance of the two language models. The second is to analyze and evaluate the translation evaluation model through experiments.

3.2 Evaluation Data Set

This paper exports the Chinese-English translation evaluated by the Language Data Consortium (LDC) on NIST Open Machine Translation 2008 as a corpus (Smt_translation), and selects 100 pairs of sentences (100 manual reference translations and 100 machine translations).

3.3 Recurrent Neural Network Translation Quality Evaluation Model

The representation of the sentence vector obtained from the recurrent neural network depends on the structure of the tree. The vector of each parent node is calculated from its directly affiliated nodes. Two child nodes c_1 and c_2 , the parent node vector is calculated by c_1 and c_2 , as shown in formula 1:

$$h_p = f(V_{Recursive} \bullet [h_{c_1}, j_{c_2}] + b_{Recursive}) \quad (1)$$

Among them $[hc_1, hc_2]$ is obtained by concatenating the vectors hc_1 and hc_2 corresponding to c_1 and c_2 , and $b_{Recursive}$ is the bias vector. f is the tangent function.

After obtaining the distributed vector representation of the sentence, a set of sentences m is constructed with a sliding window of size m . Then calculate the group score to determine whether the group is coherent, and then use the product of the coherence probabilities of all groups to determine whether the text is coherent. The calculation process is shown in formula 2:

$$p(y_{c=1}) = sig \text{ mod } (U^T * f(W_{sen} \times h_c + b_{sen}) + b) \quad (2)$$

Where $yc = 1$ means that the group is coherent, otherwise the group is not coherent. The model judges whether the text is coherent by comparing the size of the text coherence score S .

4 Analysis and Research of Machine Automatic Translation Quality Evaluation Model Based on Recurrent Neural Network Algorithm

4.1 Comparative Analysis of Scoring Results

We compare the scoring results of the QuEst method (QuEst), the scoring results of our machine automatic translation quality assessment model system based on the recurrent neural network algorithm (Our system) and the official gold standard (Gold) for comparison and analysis. Divide the sentence length in the test set into five ranges: (1,10], (10,20), (20,30], (30,40), (40, $+\infty$). Under different sentence lengths, We calculated the Pearson correlation coefficient between the QuEst system scoring results and the gold standard Gold (corresponding to the QuEst column in the table), and we also calculated the scoring results of the machine automatic translation quality evaluation model system based on the recurrent neural network algorithm established in this article and the gold standard Gold. The Pearson correlation coefficient is shown in Table 1.

Table 1. Pearson correlation coefficients of each system under different sentence lengths in the test set

Sentence length range	Number of sentences	QuEst	Our system
(1,10]	467	0.443	0.514
(10,20]	546	0.408	0.543
(20,30]	324	0.468	0.578
(30,40]	88	0.589	0.643
(40, $+\infty$)	37	-0.45	0.085

Regardless of the value range of the sentence length, the translation quality estimation system of the machine automatic translation quality evaluation model based on the recurrent neural network algorithm established in this paper is better than the QuEst system as shown in Fig. 1. We can also find that when the sentence length is greater than 40, the QuEst system and our system both obtain relatively poor score results, and the QuEst system even has a negative correlation.

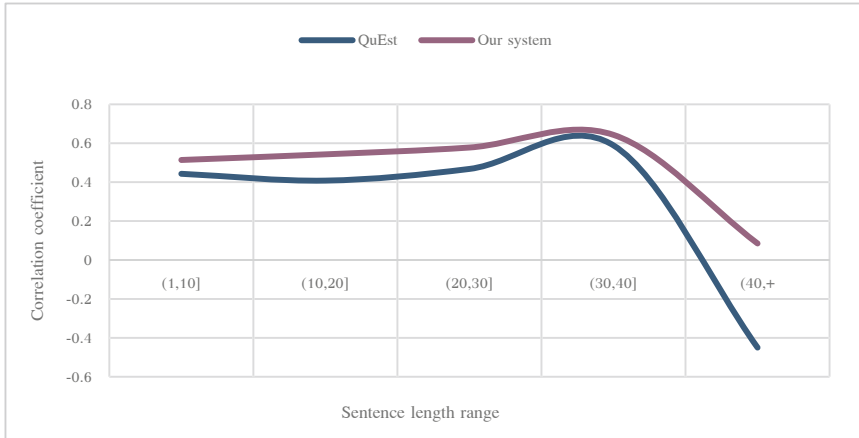


Fig. 1. Pearson correlation coefficients of each system under different sentence lengths in the test set

4.2 Language Model Performance Analysis

This paper also counts the content of nouns in the machine translation corpus, and the statistical results are shown in Fig. 2.

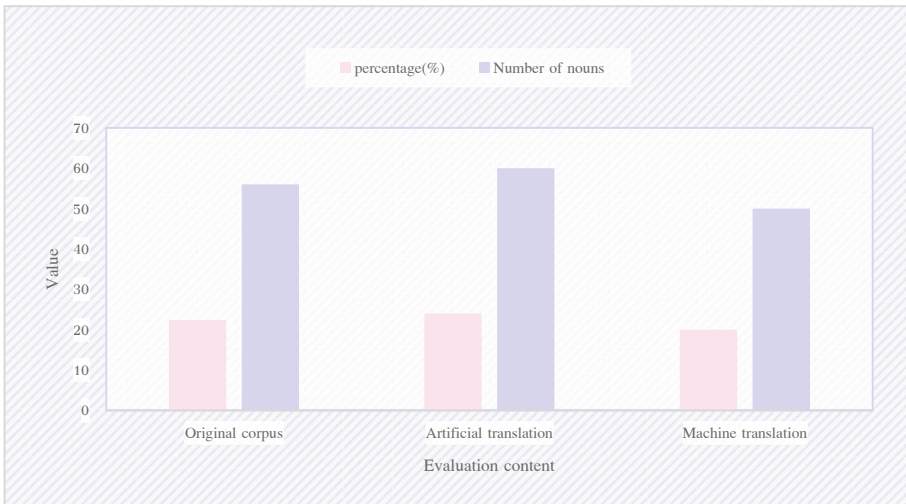


Fig. 2. Evaluate the nouns in the translation

There are about 22.4% of the nouns in the original text, about 24% of the nouns in the reference translation, and about 20% of the nouns in the translation obtained by automatic translation. Although the proportion of nouns in the text is almost the same as that of smt_translation. Chinese body ctb and composition one The body of the

series of sentences. However, because machine translation consistency detection is more difficult than sentence classification, the reason is that the former needs to distinguish between relatively similar reference translations and machine-generated translations, and the comparability ranking task of original documents and replacement documents in sentences is not high. Therefore, the recognition performance of the traditional solid grid method and the traditional neural network method is not high, but after combining the two in this article, the performance can be significantly improved. It shows that the model in this article incorporates entity information into machine learning to better evaluate the consistency of machine translation, which fully reflects the complementary role of distributed entity representation and existing deep learning models. In addition, the significance test shows that the performance of the model in Chinese is significantly improved compared with the existing models.

5 Conclusions

Language is not only the embodiment of human civilization, but also the most important communication tool for human beings. It is the main way of communication between people. People use language to preserve the crystallization of human civilization and pass it on. At the same time, with the deepening of globalization, no country can be separated from the world. Various cultures influence and infiltrate each other and promote the progress of the entire human society. This article is a research on the quality evaluation of manual translation. Taking English-Chinese vocabulary translation as an example, this paper proposes a translation quality evaluation method based on neural network language model to evaluate the quality of manual translation results from the perspective of conforming to Chinese terminology expression habits. This is a new perspective of translation quality evaluation. And further analyze the performance of the translation quality evaluation method, and finally verify the feasibility of the method through related experiments.

Acknowledgements. This work was supported by Localization of On-line Translation Courses for Non-governmental Universities in Shaan Xi under Grant No. XGH21264.

References

1. Mason, S.B., Moorkens, J., Castilho, S., Gaspari, F., Doherty, S. (eds.): Translation quality assessment: from principles to practice. *Mach. Transl.* **33**(3), 269–277 (2019)
2. Li, H., Huang, G., Cai, D., et al.: Neural machine translation with noisy lexical constraints. *IEEE/ACM Trans. Audio, Speech Lang. Process.* **8**(99), 1–5 (2020)
3. Ren, Q., Su, Y., Wu, N.: Research on Mongolian-Chinese machine translation based on the end-to-end neural network. *Int. J. Wavelets Multiresolut. Inf. Process.* **18**(1), 46–59 (2020)
4. Screen, B.: Productivity and quality when editing machine translation and translation memory outputs: an empirical analysis of English to Welsh translation. *Studia Celtica Posnaniensia* **2**(1), 1–24 (2017)
5. Kim, H., Jung, H.Y., Kwon, H., et al.: Predictor-estimator: neural quality estimation based on target word prediction for machine translation. *ACM Trans. Asian Low-Resource Lang. Inf. Process.* **17**(1), 1–22 (2017)

6. Tezcan, A., Hoste, V., Macken, L.: Estimating word-level quality of statistical machine translation output using monolingual information alone. *Nat. Lang. Eng.* **26**(1), 73–94 (2020)
7. Qiu, D., Jiang, H., Chen, S.: The integrated evaluation method of machine translation quality based on z-numbers. *IAENG Int. J. Comput. Sci.* **47**(1), 61–67 (2020)
8. Fomicheva, M., Sun, S., Yankovskaya, L., et al.: Unsupervised quality estimation for neural machine translation. *Trans. Assoc. Comput. Linguistics* **8**(1), 539–555 (2020)
9. Chatterjee, N., Gupta, S.: Efficient phrase table pruning for Hindi to English machine translation through syntactic and marker-based filtering and hybrid similarity measurement. *Nat. Lang. Eng.* **25**(1), 171–210 (2019)
10. Larroyed, A.A.: Machine translation and disclosure of patent information. *IIC: Int. Rev. Intellectual Property Competition Law*, **49**(7), 763–786 (2018)
11. Koo, S.-B.: Literary translation analysis and evaluation based on B. M. Rodríguez Rodríguez's LTQA - the case of deborah smith's the vegetarian. *J. Transl. Stud.* **18**(3), 37–64 (2017)
12. Choi, H., Cho, K., Bengio, Y.: Context-dependent word representation for neural machine translation. *Comput. Speech Lang.* **45**, 149–160 (2017)



AHP-Bayesian Network Based Analytical Model for HRMIS

Zijuan He^(✉)

School of Management, Wuhan University of Technology, Wuhan, Hubei, China

Abstract. The development of the Internet has brought opportunities and challenges to the human resource management of enterprises, and it is necessary to build human resource management information systems to maximize the effectiveness of human resource management. At the same time, the younger generation who has grown up with the reform and opening up is entering the workplace due to their own personality characteristics that are different from the previous generation, they often leave jobs frequently. In order to effectively prevent the new generation of employees from leaving the risk, taking the departure of new generation employees as an example, this paper combines AHP and Bayesian networks to provide control measures for enterprise HRM from the perspective of dynamic risk.

Keywords: New generation employees · Bayesian network · Risk

1 Introduction

With the rapid changes in the times, a new generation has entered the workplace along with the reform and development, and is called the “new generation of employees”. For this group, enterprises can be said to have mixed feelings, “happy” is that this generation of young people compared to previous employees, knowledge background and learning ability is stronger, with innovative thinking mode, willing to accept new things, for enterprise development can bring significant improvements; “worry The “worry” is that these employees are more active, more autonomous and eager to pursue results, and they find it difficult to rest on their laurels in the workplace, often leaving frequently because they cannot meet their individual needs or reach the expected standards [1].

Due to their own characteristics, “frequent job-hopping” has become a common professional characteristic of the new generation of employees [2, 3]. For companies, the frequent departure of employees, especially core employees, can bring immeasurable and irreparable losses, as it not only means higher training and management costs, but also the leakage of core secrets and loss of important customers [4]. Therefore, there is an urgent need to explore the motivation of the new generation of employees leaving the company and take corresponding measures to effectively reduce the employee turnover rate [5].

2 Characteristics of the New Generation of Employees

At this stage, as the market economy continues to develop with new vitality, the new generation of employees is gradually becoming the backbone of China's economic development, and they are changing the world in their own way, so it is necessary to have a more comprehensive and systematic understanding of employees. The new generation of employees differs from previous generations in a number of ways [6] and its characteristics are briefly summarized in the following areas.

- (1) Diversity of values. The new generation of employees shows greater inclusiveness in terms of ideas and values. For self-expectations are higher, focusing on self-centred, so when they first enter the workplace is always likely to produce a psychological gap. At the same time, this generation tends to eat the values of "realistic", with the previous demand for dedication and selflessness is different, they only accept the immediate results, more material life, good at maintaining their legitimate interests and core rights.
- (2) Strong plasticity. The knowledge level of the new generation of employees has been greatly enhanced compared to previous generations of employees, with a strong foundation of professional knowledge and the ability to operate computers proficiently, willing to accept new things and with a strong sense of innovation. Driven by strong motivation, they are often able to actively learn and continuously improve their overall quality. The company is also able to make them recognize the values of the organization and maintain a high level of enthusiasm for innovation through the inculcation of the organizational culture.
- (3) Mental quality is not strong. Although the new generation of employees are brave enough to express their inner thoughts, but due to their birth stage and period of relatively rich material life, growing up stage are dependent on the care and love of parents, once into the workplace, away from the natural shelter of the family, the new generation of employees inevitably show some more childish behavior and ideas. The ability to solve complex problems needs to be improved, and negative emotions are difficult to resolve in a timely manner, which can easily lead to extremes.
- (4) Changing career outlook. Young employees attach more importance to the quality of life and are prone to emotions in the process of facing complex problems at work. As their subjective feelings largely influence their decisions, subjective well-being can easily affect the job performance of new generation employees. Many employees are emotionally restless, unwilling to work solidly and steadily, empty of slogans, and difficult to give consistently in a position, so when faced with higher salary levels, they often choose to separation.

In the face of the significant characteristics of the young employees, and in-depth understanding of their needs, to find their own strengths and abilities to play the motivation, set the corresponding incentive mechanism, so as to reduce the rate of employee turnover, improve the effectiveness of corporate innovation capacity.

3 Establishment of a Risk Assessment Index System

3.1 Influencing Factors of New Generation Employees Leaving

The new generation of employees has received a lot of attention from scholars at home and abroad, and especially the factors influencing the departure of new employees have been explored more deeply. From the perspective of organizational fairness, Han Hongjun et al. mainly explore the influence mechanism of distributive fairness and procedural fairness on the tendency of new generation employees to leave and its effect. In addition, there are also scholars based on the perspective of organizational culture. The concept of “cultural retention” is proposed, and it is believed that the error-averse culture has a positive impact on the tendency of new generation employees to leave the company [4].

Through a review of domestic and foreign scholars’ research [4, 3, 7], and combined with the personality characteristics of new generation employees, the factors that affect the tendency of employees to leave are roughly divided into the following four categories. They are as follows:

- (1) Personal factors: work-family conflict, breakdown of psychological contract, and inability to realise personal values.
- (2) Job factors: low job satisfaction, low job involvement, job burnout, poor pay incentives, lack of career development guidance, and low sense of job achievement.
- (3) Organizational factors: unreasonable reward system, inadequate promotion mechanism, poor organizational climate, error-averse culture, low sense of organizational fairness, poor company effectiveness and prospects.
- (4) Environmental factors: high cost of living pressure, work opportunities outside the organization.

3.2 Employees’ Exit Risk Assessment Index System

This paper proposes to establish a new generation of employees’ exit risk assessment index system, which mainly contains qualitative and quantitative indicators, i.e. static and dynamic indicators. Static indicators are mainly obtained through expert interviews and questionnaires, and the data analysis part will be carried out through Yaahp hierarchical analysis software after scoring by experts to further determine quantitative indicators and indicator weights; dynamic indicators are mainly delivered in real time with the help of the enterprise content’s human resource intelligence system (employees’ weekly summaries and feedback on the enterprise), and the data collected at the enterprise end will be analyzed and collated through cloud computing and big data technology. The data collected at the enterprise end is collated through cloud computing and big data technology for text data analysis, and the processed data can be directly substituted into the calculation to achieve intelligent assessment of employee departure risk.

For example, the assessment indicators of organizational factors can be extracted and calculated through textual analysis of visual data such as C11 unreasonable reward system, C12 inadequate promotion mechanism, C13 poor organizational atmosphere, C14 error aversion culture and C15 low sense of organizational fairness. This paper combines both dynamic and static aspects to refine the four assessment indicators in the guideline layer into 17 quantitative first-level indicators. A quantifiable and intelligent risk assessment indicator system is constructed (Fig. 1).

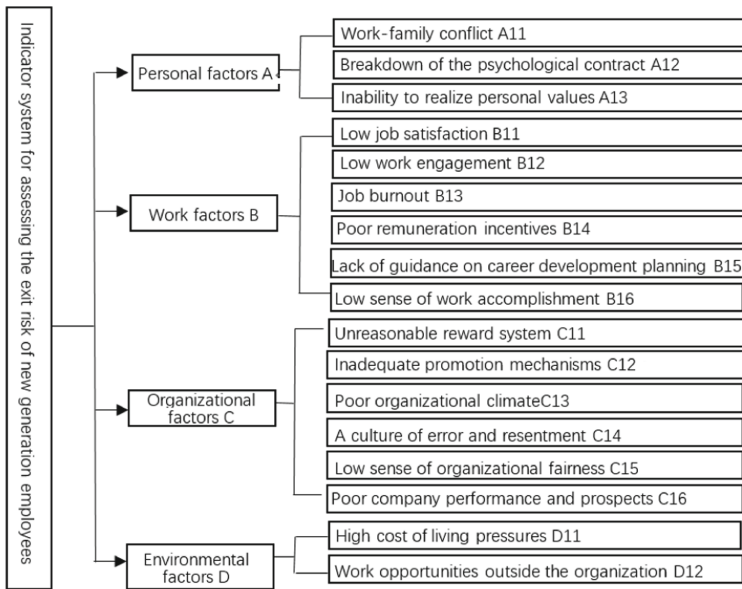


Fig. 1. New generation employee exit risk assessment indicator system

4 Construction of the Separation Risk Assessment Model for New Generation Employees

4.1 Subjective Weighting Based on the AHP Method

The basic principle is to divide a complex assessment system into four levels, such as objectives, criteria, sub-criteria and indicators, analyse the interactions of the influencing factors, quantify them by means of a specified scale, construct a judgment matrix and calculate the weights of the influencing factors by calculating the maximum eigenvalues of the judgment matrix. The specific analysis steps are shown below.

Hierarchical analysis is a hierarchical weighting decision analysis method based on multi-objective comprehensive evaluation, decomposing the key problem to be solved into a number of levels such as objectives, effects and indicators, combining quantitative and qualitative aspects to provide a more convenient decision optimisation models for

complex decision problems with multiple objectives, multiple criteria and no significant structural features [8]. The specific application is carried out in four major steps: (1) construction of a structural model (2) construction of pairwise comparison matrices (3) calculation of weight vectors and consistency testing (4) calculation of combined weight vectors and combined consistency testing.

4.2 Objective Weighting Based on Bayesian Network Method

The process of constructing a Bayesian network mainly includes structural learning and parametric learning, which consists of two major steps: firstly, the correlation between the variables is calculated and some of the less relevant variables are eliminated to improve the efficiency and correctness of structural learning, secondly, the causal relationship between nodes and nodes is determined by the structural learning method of the K2 algorithm, and finally, the nodes are determined by parametric learning. The posterior probability distribution is used to quantify the node correlation through large sample data. This paper adopts the Bayesian network method that combines the data obtained from sample collection and prior experience to carry out parameter learning and analyse its validity on the basis of constructing a model [9].

In order to understand in real time the departure trends of the new generation of employees, this paper is devoted to the study of dynamic Bayesian networks to dynamically rate and predict the departure risk. The dynamic Bayesian network consists of two main components: (i) an initial state, Bayesian network N0; and (ii) a Bayesian network N of two or more time segments (Table 1).

Table 1. Example of a dynamic Bayesian network

	X-0	X1-0	X2-0
State0	50.0	50.0	50.0
State1	50.0	50.0	50.0

The probability calculation using Bayesian networks needs to follow the following principles: with reference to Bayesian networks, probability integrals are used to describe uncertainty, and for probability integrals, the probability of the associated event needs to be taken as follows.

$$P(A, B) = P(A/B) \times P(B) \tag{1}$$

In the above equation, P(A, B) denotes the actual probability of events A and B occurring, P(B) denotes the probability of event B occurring, and P(A/B) denotes the probability of event A occurring conditional on event B. P(A/B) can be converted into Eq. (2).

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)} \tag{2}$$

If once more than one data is involved, the Bayesian rule needs to be extended, e^- is assumed to be the value of the vector obtained by the n sensors and M_i is assumed to be the set of H states of M .

$$P(M_i/e^-) = \frac{P(e^-/M_i)P(M_i)}{\sum_{i=1}^H P(e^-/M_i)P(M_i)} \tag{3}$$

Dynamic Bayesian networks can achieve dynamic and intelligent prediction of new generation employees' departure, and can play the role of early warning and control of departure [10] (Fig. 2).

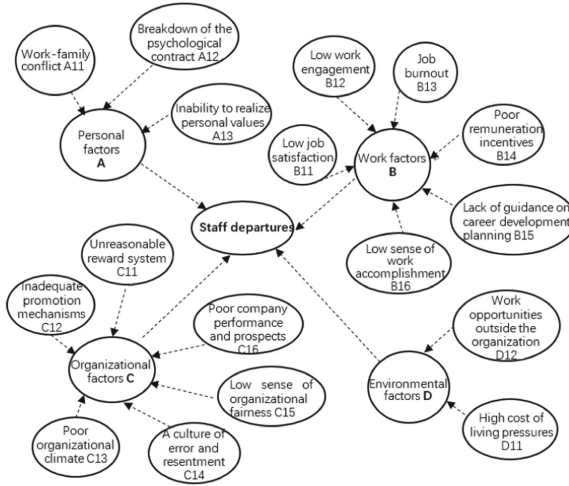


Fig. 2. Bayesian network diagram for the assessment of exit risk for new generation employees

4.3 Construction of an AHP-Bayes-Based Risk Assessment Model

In this paper, we use AHP and Bayes network to assess the risk of employee turnover. The main assessment process is shown in the above diagram (Fig. 3): the assessment index system is divided into three levels, such as target, criterion and sub-criterion. The AHP method is effectively coupled with the Bayes network method to determine the reasonable weights of the assessment indicators, on the basis of which the maximum weighting principle is used to evaluate the risk level of separation, and different risk countermeasures are taken at different risk levels [11].

According to the results of the software, it is found that the personal factors of inability to realise personal value, job-related burnout and low work engagement and poor pay incentives, organizational factors of low organizational fairness and extra-organizational work opportunities become the key factors for employees to leave, which is also closely related to the personality characteristics of the new generation employees [12], whose personality of pursuing freedom and fairness will lead to the motivation

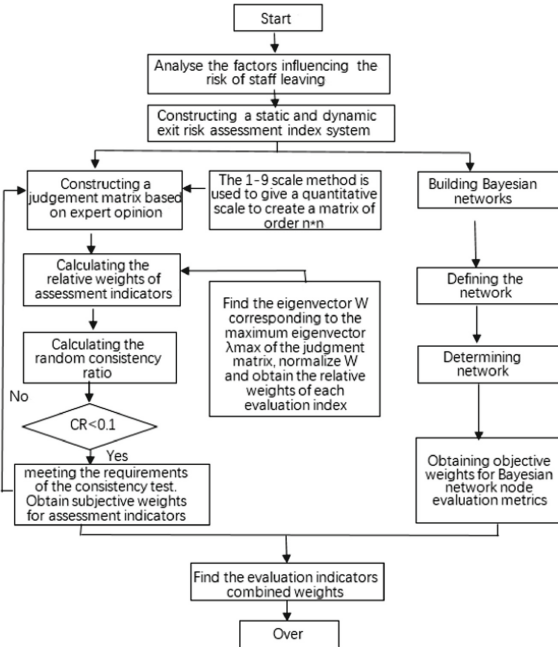


Fig. 3. Flowchart for assessing the risk of new generation employees leaving the company

to leave due to the combination of various factors. This is also closely related to the personality traits of the new generation of employees, whose pursuit of freedom and a sense of fairness will lead to the motivation to leave the company and thus to the act of leaving, which provides a favourable reference for companies to adopt intervention measures for the new generation of employees to leave.

5 Conclusion

The new generation of staff has played a pivotal role in the development of the enterprise, administrators should start from the actual situation, the key factors affecting the departure, to take targeted measures.

- (1) Clear salary incentive mechanism. The new generation of employees pursuing material benefits, as the economic basis of their lives, it is necessary to explore ways to reduce staff turnover from the perspective of pay incentives, while pay and corporate costs want to link, this is an important issue that companies must weigh and consider.
- (2) Clarify employees' career plans. The new generation of employees advocates freedom, enterprises can learn from the management experience of some excellent enterprises, such as the use of "passing on help", "mentorship", "mentorship" and "mentoring". "Mentorship", "coaching" and other career planning methods, so that

the old staff to help new employees, so that new employees to stimulate the enthusiasm of the old staff to work, so as to achieve the purpose of mutual promotion and common development.

- (3) Clear promotion channels. For new employees, while pursuing material benefits, they also attach importance to their spiritual needs, especially for promotion. New generation employees are open-minded, autonomous and goal-oriented, so they dare to challenge authority and achieve their own goals, and often stand out in the goal assessment, so reasonable promotion channels undoubtedly play a stronger motivational role.

References

- Jennings, C.R.: Social and economic characteristics as determinants of residential fire risk in urban neighborhoods: a review of the literature. *Fire Saf. J.* **62**, 13–19 (2013)
- Harris, K.J., Wheeler, A.R., Kacmar, K.M.: Leader–member exchange and empowerment: direct and interactive effects on job satisfaction, turnover intentions, and performance. *Leadersh. Q.* **20**(3), 371–382 (2009)
- Cenk, S., Toz, A.C., Buber, M., Koseoglu, B.: Risk analysis of grounding accidents by mapping a fault tree into a Bayesian network. *Appl. Ocean Res.* **113**, 102764 (2021)
- Koseoglu Balta, G.C., Dikmen, I., Birgonul, M.T.: Bayesian network based decision support for predicting and mitigating delay risk in TBM tunnel projects. *Autom. Constr.* **129**, 103819 (2021)
- Williams, G.: Management millennialism: designing the new generation of employee. *Work Employ Soc.* **34**(3), 371–387 (2019)
- Karade, M.M., Gankar, S., Bhagwat, A.: A perception analysis of managers and gen Y employees towards job opportunities. *J. Indian Manag. Res. Pract.* **1**(1), 14–24 (2015)
- Kendall-Raynor, P.: NHS plan at risk from staff shortages. *Nurs. Stand.* **33**(9), 7 (2018)
- Emek, Y., Kutten, S., Lavi, R., Shi, Y.: Bayesian generalized network design. *Theor. Comput. Sci.* **841**, 167–185 (2020)
- Kanes, R., Ramirez-Marengo, C., Abdel-Moati, H., Cranefield, J., Véchet, L.: Developing a framework for dynamic risk assessment using bayesian networks and reliability data. *J. Loss Prev. Process Ind.* **50**, 142–153 (2017)
- Sunder, S., Kumar, V., Goreczny, A., Maurer, T.: Why do salespeople quit? an empirical examination of own and peer effects on salesperson turnover behavior. *J. Mark. Res.* **54**(3) (2017)
- Yu, J., Tian, S., Wang, W., Jiang, Z., Yang, J., Wei, S.: A dynamic intelligent fire risk assessment model based on AHP-Bayes for urban old neighborhoods - an example of M neighborhood in Shanghai. *Saf. Environ. Eng.* **28**(05), 10–17+50 (2021). (in Chinese)
- Wang, J.H., Hao, W., Tao, Z., Wei, L.: Safety risk assessment of tower crane operations based on fuzzy Bayesian networks. *Saf. Environ. Eng.* **28**(04), 15–20 (2021). (in Chinese)



Medical CT Image Enhancement System Based on Swarm Intelligence Optimization Algorithm

Li Xu¹, Yuqing Hou², Fengjun Zhao², and Jinniu Bai¹(✉)

¹ Department of Computer Science and Technology, Baotou Medical College, Baotou 014040, China

² School of Information Science and Technology, Northwest University, Xi'an 710127, China

Abstract. With the development of electronic technology and computer technology, the application field of medical images will continue to expand, and medical image processing technology will also attract more and more attention. CT images are increasingly used for clinical diagnosis and treatment. The purpose of this work is to study the CT image enhancement system based on swarm intelligence optimization algorithm. Integrate medical image enhancement algorithms based on swarm intelligence optimization algorithms and common medical image processing and analysis algorithms into the medical image processing system to form a powerful and easy-to-use medical image enhancement system. Based on the research of adaptive gray-scale image enhancement technology based on intelligent optimization algorithm, a new target fitness function is proposed, which uses GA, PSO and QPSO to achieve adaptive gray-scale image enhancement. The results show that the suitability value of the QPSO group intelligent optimization algorithm is 9.5241, the robustness is strong, and the enhancement effect is slightly better.

Keywords: Swarm intelligence optimization · Medical CT · Image enhancement · Medical system

1 Introduction

Image enhancement is one of the most basic methods of digital image processing [1]. This has two main purposes: one is to improve the visual effect of the image and the sanctity of the image elements, and the other is to make the image more suitable for human observation or automatic computer analysis [2]. In the early stages of the image, improving the image is a very important issue and a valuable technique. Image enhancement is an extensive research field in the field of imaging, and in-depth research is very important for the further application of medical CT scans in diagnosis and treatment [3, 4]. At present, swarm intelligence optimization algorithms and image enhancement technologies are both very active research areas, and intelligent optimization algorithms and image enhancement technologies and their applications have very important values [5].

The performance of medical imaging technology, especially CT monitoring, is often affected by the poor contrast quality detected by some medical imaging equipment

[6]. We recommend using the contrast enhancement method as a solution to adjust the intensity distribution of the black image. Karel F proposed advanced digital imaging and simple black medical imaging. The method is based on automatic gamma correction using different wave conversion and single value (DWT-SVD). In the first step, the technology uses DWT to transmit incoming medical images to four frequency bands, and then calculates a single value for the low-bandwidth sub-band (LL). In the second step, use sufficient reproducibility factors and single dose (SVD) factors to develop advanced LL components. In the third step, in order to further advance the LL component, the LL subband image obtained in the SVD enhancement stage is divided into two categories (small contrast and medium contrast) according to its statistical information, so we use gamma fencing. Used to process correction work. In fact, the automatic gamma correction factor is calculated based on the shape of each image. Finally, the sub-band LL image obtained by DWT conversion is combined with low intensity (LH), low resolution (HL) and high resolution (HH) to enhance image vision. Different types of non-invasive CT image plans are used for performance evaluation of the proposed contrast enhancement algorithm based on automatic gamma correction using DWT-SVD (DWT-SVD-AGC). The results show that the performance of the proposed algorithm is better than other state-of-the-art technologies [7]. Slavine N uses 4D XCAT Phantom to simulate small and small contrast lesions implanted in the liver. In all Phantom studies, the image is guaranteed to have high resolution and low noise compared to FBP. Generally speaking, repetitive motion and recovery reach the highest SNR and CNR values after about 20 repetitions, and the progressive coefficient of CNR and SNR in noisy CT images is about 1.5 [8]. Research on medical CT image enhancement system based on swarm intelligence optimization algorithm has practical significance.

This article conducts a comprehensive research on the commonly used medical image processing methods for CT images, applies image processing theory, improves CT images, and discusses and analyzes the experimental results. The information contained in CT images is extremely complex, and it is inevitable that there will be incomplete and inaccurate problems. Custom image enhancement using QPSO Group's intelligent optimization algorithm performs CT image enhancement. In the process of CT image processing, the adaptive image enhancement algorithm of the QPSO team's intelligent optimization algorithm is introduced, and optimization methods such as genetic algorithm (GA), particle swarm algorithm (PSO) and quantum swarm algorithm are proposed. The analysis of image enhancement results proves that the quantum behavioral particle swarm intelligent optimization algorithm has a good image enhancement effect.

2 Research on Medical CT Image Enhancement System Based on Swarm Intelligence Optimization Algorithm

2.1 Intelligent Optimization Algorithm of Particle Swarm Based on Quantum Behavior

The QPSO algorithm based on quantum behavior is improved on the basis of the classical particle swarm optimization (PSO algorithm). It mainly combines the ideas of quantum physics and modifies the "evolution" method of PSO (the installation method, which is

mainly the local information of the current optimal crystal position of each particle and the optimal position of global interference) account merger [9, 10].

Compared with the traditional PSO algorithm, the QPSO algorithm has the following advantages: it improves the convergence of global optimization, is unlikely to fall into local optimization, and has powerful global search, local search and global search capabilities. Although the algorithm has weak local search ability. The local search ability is relatively strong. The convergence speed is fast, and the QPSO algorithm can quickly converge to the global optimum [11, 12].

2.2 Image Enhancement

The following is the method of image enhancement: Image enhancement technology can be divided into field enhancement method and frequency correction method according to the different areas where the improvement process occurs. People must use direct images to formulate a two-dimensional space, that is, directly visualize the grayscale size of each gift. The second is to first convert the image of the field area into a frequency conversion model according to certain rules. Then edit the image in the frequency field, and then switch to the field complex.

Space-based optimization methods can be divided into two modes: gray conversion and site optimization according to different technologies. Gray transition is a gradual process based on on-site activities. The gray value of each node is converted into a new gray value according to certain types of mathematical variables, such as the number of contrast enhancement levels and common histogram equations. Website creation is a gradual process based on local space. Use the template to perform specific mathematical tasks for each gift and each gift in the local area to obtain the new gray value and the maximum number of pixels generated. This is not only related to the gray value of the gift, but also related to the gray value of the neighboring pixels. Conventional image enhancement and capture functions are real-time tasks.

2.3 The Architecture of the Image Enhancement System

(1) System structure

Image enhancement unit: Because the captured image is blurred, it needs to be enhanced. The program will be designed to process images using methods based on partial variance equations. Data management: The main purpose is to complete the record of patient information, the efficient processing of images according to the image enhancement program, the proposition of attribute values, and the storage of structured attribute data in the database for retrieval. Diagnostic evaluation: The main purpose of measuring the size of reference samples and retrieving images is to improve the accuracy of diagnosis according to the needs of the examiner.

(2) Software design

Image acquisition and input unit: adopts a variety of image acquisition methods, supports a variety of standard and digital non-standard or analog video projection formats. Equipped with standard DICOM image acquisition and non-standard DICOM, video scanning, film, etc. Image editing unit: Provide basic image editing system: image enhancement, feature extraction, editing, etc. The ultimate goal

of image processing is to allow researchers to easily analyze the content of ultrasound images and pinpoint the location of scratches. Pathology Assistant: Provide image information, measurement and statistical analysis to promote continuous improvement of clinical diagnosis.

3 Research and Investigation of Medical CT Image Enhancement System Based on Swarm Intelligence Optimization Algorithm

3.1 Simulation Experiment

This article introduces the development and testing of medical imaging systems on computers with commonly used configurations (Intel P42.4G processor, 256M memory, 64M graphics card). The monitor uses the HRCT (High Quality CT) breast of a 55-year-old woman who was admitted to the Health Hospital of the Medical College of M Province on October 10, 2021 as an experiment to determine a specific medical imaging operating system. The model is improved based on the lung tissue in the human heart HRCT. The model is a clinical imaging function based on the QPSO group intelligent optimization algorithm created by the system.

Using the degraded grayscale image frog, the size is 256×256 , and each algorithm system runs 30 times. In the implementation of the GA algorithm, the population size is set to 60, the crossover rate is set to 0.6, the conversion probability is initially set to 0.01, the chromosome is a 20-bit binary bit, the number of evolutionary variables is 500, and the population size is set in the PSO and OPSO algorithms 30. The number of evolutionary iterations is 500.

3.2 Image Adaptive Enhancement Process of QPSO Group Intelligent Optimization Algorithm

(1) Initialization

Set the group size, that is, the number of particles, here is set to 30;

Initialize the position vector of each particle of the population;

Initialize the local optimal position for any f : $Pbest(i) = x_i$;

Initialize the global optimal position $Gbest$ as: $Gbest = \max\{f(x_1), f(x_2), \dots, f(x_N)\}$;

(2) Calculate the objective function value of each particle in the particle swarm

In the simulation example, the objective function we use is shown in formula 1:

$$Fitness(i) = \frac{1}{n} \sum_{x=1}^M \sum_{y=1}^N f'^2(X, Y) - \left[\frac{1}{n} \sum_{x=1}^M \sum_{y=1}^N f'(x, y) \right]^2 \quad (1)$$

$$+ \sum_{x=1}^{M-2} \sum_{y=1}^N (f'(x, y) - f'(x+2, y))^2$$

Among them, M and N represent the length and size of the image, $n = M \times N$, $f'(x, Y)$ represents the gray value of the original image resolution (X, Y) after general conversion, and represents the i -th particle. The larger the amount of $Fitness(i)$, the better the image enhancement effect.

For each particle, compare your fitness value with the fitness value of the best $Pbest$ position you have experienced. If it is better, use the current position x_i as the new $pbest$. Fitness is a field activity that needs to be increased. The current state of the best atom of particle i is shown in Eq. 2:

$$Pbest_i(t + 1) = \begin{cases} Pbest_i(t) & f(X_i(t + 1)) \leq f(Pbest_i(t)) \\ X_i(t + 1) & f(X_i(t + 1)) > f(Pbest_i(t)) \end{cases} \quad (2)$$

Among them, the subscript “ i ” represents the i -th particle, and t represents the number of iterations.

4 Research and Analysis of Medical CT Image Enhancement System Based on Swarm Intelligence Optimization Algorithm

4.1 Comparison of the Effects of Group Intelligence Optimization Algorithms

The fitness function obtained using the QPSO algorithm is slightly larger than the GA and PSO algorithms, as shown in Table 1. The fitness function obtained by using the PSO algorithm is slightly larger than that of the GA. It can be seen that the enhancement effect of the QPSO algorithm in the gray-scale image enhancement algorithm is slightly better than that of the GA and PSO algorithms, and the enhancement effect of the PSO algorithm is slightly better than that of the GA.

Table 1. Comparison of α , β and fitness in the simulation results of GA, PSO and QPSO algorithms

Algorithm	α	β	Fitness
GA	9.2564	9.0214	8.5245
PSO	9.5245	9.6542	8.6511
QPSO	9.9895	9.9524	9.5241

The comparison of the mean and standard deviation in the simulation results of adaptive gray-scale image enhancement of GA, PSO and QPSO algorithms is shown in Fig. 1. Larger, followed by PSO, and slightly smaller than GA. However, the standard deviation of QPSO is slightly larger than that of PSO, the standard deviation of GA is the smallest, and PSO is somewhere in between. OPSO is good.

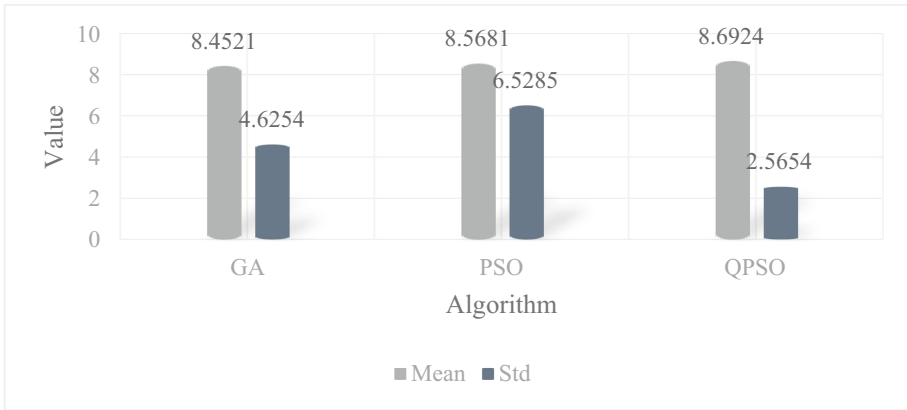


Fig. 1. Comparison of mean and standard deviation in simulation results of GA, PSO and QPSO algorithms

4.2 Realization of Medical Image Enhancement System

The medical image enhancement system for image adaptive enhancement based on the QPSO group intelligent optimization algorithm proposed in this paper realizes the enhancement function for the target area. Contrast of medical images in the area. Use a variety of image enhancement algorithms to enhance and compare the target area of medical images to find the most effective upgrade enhancement algorithm.

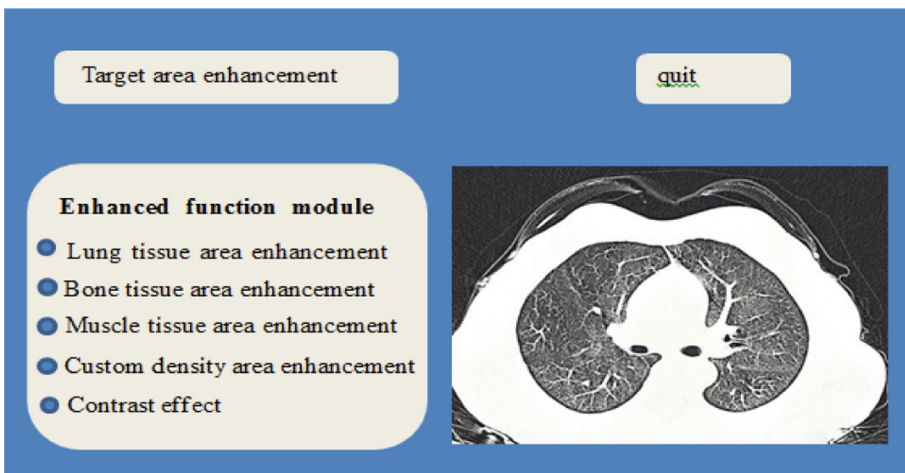


Fig. 2. Enhancement of different target areas

The statistical algorithm-based imaging used in this program provides “technical” clinical images (weight differences between human bodies) to determine the normal distribution points and distribution of specific HRCT medical imaging materials in the

area. Enhance and modify the sub-image and background image of the region, and merge to obtain an enhanced image. Taking the lung area of the lung tissue on the HRCT chest image will increase the focus of the lung area in the image. We have performed many tests on medical images, compared the QPSO team intelligent optimization algorithm used in this system with some classic improved algorithms, and conducted research on DSM (Probability Assignment Evaluation Method). Therefore, it is ensured that the QPSO group intelligent optimization algorithm used in this system is an improved algorithm that is more suitable for medical image enhancement.

5 Conclusions

Medical image enhancement is a hot topic that is being widely studied in the field of image processing. Its in-depth research is of great significance for the further application of medical CT images in clinical diagnosis and treatment. The traditional optimization algorithm and modern intelligent optimization algorithm in image enhancement are studied, and the image enhancement algorithm optimized by quantum particle swarm optimization algorithm is proposed. Our optimization methods such as genetic algorithm (GA), particle swarm optimization (PSO), and quantum particle swarm optimization (QPSO) perform enhancement experiments on CT images. The experimental results show the robustness and accuracy of the registration results of intelligent optimization algorithms such as PSO and QPSO All are better than the genetic algorithm optimization method; the QPSO quantum particle swarm optimization algorithm can solve the problem of the optimal solution in the space.

Acknowledgements. This work was supported by Inner Mongolia Natural Science Foundation of China under Grant No. 2019MS06036.

References

1. Mo, J., Zhang, L., Wang, Y., Huang, H.: Iterative 3D feature enhancement network for pancreas segmentation from CT images. *Neural Comput. Appl.* **32**(16), 12535–12546 (2020). <https://doi.org/10.1007/s00521-020-04710-3>
2. Rong, Z., Chu, C., Zhang, Y., et al.: Cultural relic image enhancement based on the Laplacian of the gaussian and retinex model. *J. Comput. Theor. Nanosci.* **14**(8), 3692–3697 (2017)
3. Uhlig, J., Fischer, U., von Fintel, E., et al.: Contrast enhancement on cone-beam breast-CT for discrimination of breast cancer immunohistochemical subtypes. *Translat. Oncol.* **10**(6), 904–910 (2017)
4. Brown, J., et al.: X-ray CT in phase contrast enhancement geometry of alginate microbeads in a whole-animal model. *Ann. Biomed. Eng.* **48**(3), 1016–1024 (2019). <https://doi.org/10.1007/s10439-019-02291-4>
5. Wu, C., Liu, Z., Jiang, H.: Choosing the filter for catenary image enhancement method based on the non-subsampled contourlet transform. *Rev. Sci. Instrum.* **88**(5), 453–476 (2017)
6. Tang, H., Pei, X., Huang, S., et al. Automatic lumbar spinal CT image segmentation with a dual densely connected U-Net. *IEEE Access* (99), 1 (2020)
7. Kallel, F., Hamida, A.B.: A new adaptive gamma correction based algorithm using DWT-SVD for non-contrast CT image enhancement. *IEEE Trans. Nanobiosci.* (99), 1 (2017)

8. Slavine, N., Mccoll, R., Oz, O., et al.: Phantom and pre-clinical studies for image improvement in clinical CT. *IEEE Trans. Radiat. Plasma Med. Sci.* **3**(1), 96–102 (2019)
9. Xu, X., Zhou, F., Liu, B., et al.: Multiple organ localization in CT image using triple-branch fully convolutional networks. *IEEE Access* (99), 1 (2019)
10. Achenbach, S., Paul, J.F., Laurent, F., et al.: Comparative assessment of image quality for coronary CT angiography with Iobitridol and two contrast agents with higher iodine concentrations: Iopromide and Iomeprol. a multicentre randomized double-blind trial. *Eur. Radiol.* **27**(2), 821–830 (2017)
11. Yu, H., Ge, W.: Sart-type half-threshold filtering approach for CT reconstruction. *IEEE Access* **2**(2), 602–613 (2017)
12. He, X., Nie, Y., Guo, H., et al.: Research on a novel combination system on the basis of deep learning and swarm intelligence optimization algorithm for wind speed forecasting. *IEEE Access* (99), 1 (2020)



Design of English Multimedia Intelligent Translation System Based on Neural Network Algorithm

Zheng Zhao^(✉)

Xi'an Fanyi University, Xi'an 710000, Shaanxi, China
zhaozheng@graduate.utm.my

Abstract. Due to the vigorous development of network information technology and the internationalization of the market economy, communication in various fields is becoming more and more common, and the requirements for cross-language communication are increasing. As a simple and efficient tool, the machine translation system can achieve peer-to-peer translation between various sentences while maintaining the original semantic structure, which is very critical in practice. This paper aims to study the design of English multimedia intelligent translation system based on neural network algorithm. Based on the analysis of common neural network translation models, neural network learning rules and system non-functional requirements, the English multimedia intelligent translation system level is designed. In order to test the translation performance of the designed system, the system in this paper is compared with other systems. The test results show that the performance of the translation system in this paper is optimal regardless of the length of the interval, and as the sentence length gradually increases, the translation performance gradually increases.

Keywords: Neural network algorithm · Machine translation · English translation system · Multimedia intelligence

1 Introduction

With the improvement of computer technology and the rapid development of artificial intelligence technology, the pace of informatization has increasingly shifted to other countries, and the requirements for translation have also increased significantly. Translation has become the main method of communication between various voices, and its significance is self-evident [1, 2]. Traditional manual translation is slow and costly, and is not suitable for large-scale text translation. In addition, the translation results vary from person to person, and the quality of translation is not reliable, and it is gradually unable to adapt to people's actual needs [3, 4].

In the 21st century, due to the development and maturity of machine translation methods, many models try to enter the market. Some researchers extract aligned segments from a pair of aligned Chinese and English bilingual parallel sentences, and statistical machines use different levels to compile the Chinese and English versions, and launch a

multi-level filtering algorithm based on automatic compilation [5, 6]. Some researchers have also used language technologies ranging from text to speech and language levels to establish a statistical machine-based translation model and completed a rare translation from language to Chinese [7, 8]. In the context of deep learning neural network computing and word vector technology in natural language processing, some researchers have provided a neural network-based machine translation model and established an end-to-end architecture to provide and Modeling, End-to-End architecture will also serve as the standard architecture of the future neural machine translation system. The design innovations of these schemes overturned the linear modeling of traditional calculation methods in machine translation, integrated translation modes, alignment models, language models, etc. into statistical methods, and improved translation capabilities through the use of neural network capabilities [9, 10]. Some researchers have improved the alignment model based on neural machine translation to detect the alignment of bilingual words and improve the translation quality [11, 12]. Neural network learns new representations of various objects in natural language, and proposes new ideas to solve the problems of sparse features and lack of ability to describe language structure in natural language processing.

On the basis of consulting a large number of relevant references, this paper combines common neural network translation models, neural network learning rules and system non-functional requirements, designs an English multimedia intelligent translation system, and compares this system with the other two systems to test whether the system performance is good when translating sentences of different lengths.

2 Design of English Multimedia Intelligent Translation System Based on Neural Network Algorithm

2.1 Common Neural Network Translation Models

(1) Translation model based on recurrent neural network

A common feature of translation models based on recurrent neural networks is that the encoder and decoder elements are both recurrent neural networks. Common variants can be divided into two aspects: layer stacking and directional stacking. Generally speaking, deep networks are more powerful than single-layer networks in understanding and expression, but increasing the number of network layers will make training more difficult and require more training data. Directional superposition refers to the use of two cyclic neural networks to encode sequence data from two different directions, forward and reverse, and combine the two results to form the final result, which can improve the model's ability to understand sequence data.

The time-expanded RNN calculation method is inherently suitable for processing sorted data such as translated text. The general RNN calculation process is given by the following formula.

$$s_t = f(Ux_t + Ws_{t-1} + b) \quad (1)$$

s_t represents the hidden state of the recurrent neural network at time t .

The calculation process of the final output y_t at time t is shown in Eq. (2), where g is also a nonlinear activation function.

$$y_t = g(Vs_t + b) \tag{2}$$

(2) Translation model based on convolutional neural network

Different from the calculation of the recurrent neural network, the convolutional neural network needs the calculation result of the previous moment. Convolutional neural networks can well support parallel functions. Therefore, in terms of model efficiency, the neural machine translation model based on CNN is significantly better than the neural machine translation model based on RNN. Similarly, the CNN-based model does not have the problem of slope loss or slope burst caused by too large sequence, but it will continue to occur as the number of network layers increases. Figure 1 shows the schematic structure of this type of model.

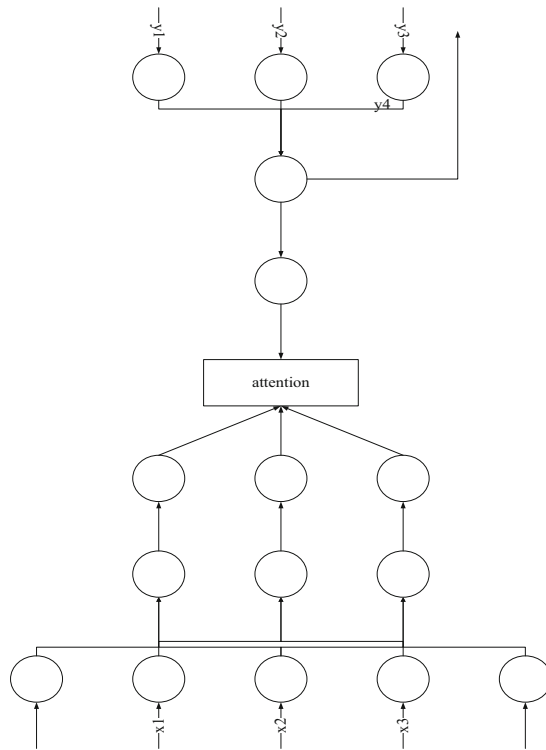


Fig. 1. CNN-based neural machine translation model

(3) Translation model based on self-attention mechanism

Unlike a convolutional neural network that captures the dependency relationship between words window by window, as the number of layers in the convolutional network increases, the continuously abstract dependency information is combined with the

dependency information of the entire sentence. The self-attention mechanism module is used as a component of sequence function mining. Although it can achieve a high degree of parallel computing, the low-level network can capture the dependency information between each word pair in the sequence. In addition, the self-care mechanism does not need to consider the distance between two words when capturing the relationship between two words, which is better than the RNN-based translation model.

2.2 The Learning Rules of Neural Networks

Neural networks usually complete specific learning tasks through iterative training. However, for different task purposes, the learning rules are not the same. They are usually divided into the following three categories.

(1) Supervised learning

In supervised learning, the learning rules are given by a set of examples (training set) that explain the characteristics of the network. Such as instance set $\{(p_1, t_1), (p_2, t_2), \dots, (p_Q, t_Q)\}$, where p_Q is the input of the network, and t_Q is the corresponding target output. When input p_Q acts on the network, the network calculates the error between the actual output and the target output, adjusts the network weight and bias value according to the set learning rules, and finally the actual network output is infinite, reaching the target and close to the target output.

(2) Unsupervised learning

In unsupervised learning, there is no target output, and the network weight and bias value are only adjusted according to the network input. The goal of this type of algorithm is usually similar to grouping. In other words, network learning divides the input into a limited number of types. Supervised learning should be based on a large number of labeled training samples composed of input signals and corresponding expectations. However, in practice, collecting labeled samples is a time-consuming and labor-intensive task, and many experiments have the problem of insufficient samples. However, unsupervised learning only requires samples, and these samples do not require labels. The sample size is sufficient because they only contain one set of input signals. Considering this situation, semi-supervised learning is produced, and the training samples are composed of a mixture of labeled and unlabeled samples.

(3) Reinforcement learning

Reinforcement learning is somewhere between supervised and unsupervised learning. It is based on the continuous interaction between the learning system and the environment. This is different from supervised learning, which provides a response target for each item, but only one evaluation level. This scoring level is a measure of the network performance of a specific input sequence. At present, this kind of learning method is relatively rare, and is generally suitable for control system applications.

2.3 System Non-functional Requirements

(1) Stability

In the Internet era where user experience is paramount, stability is a key performance. If you want to ensure the stability of users and get praise from users, stability is the most

important thing. During the operation of the system, serious errors should be avoided and the possibility of system errors should be minimized.

(2) Fluency

The fluency of the system is a very important performance. Many optimizations have been made at the bottom of the system to facilitate application. The underlying support can significantly guarantee the progress of the system. In addition, for asynchronous high-resource loads, multi-threading technology should be used to facilitate user interaction and minimize the impact of latency. Pay attention to interface design, reduce redundant data, reduce the downtime of users with poor network conditions, increase enough user prompts, and improve the fluency of the system when network conditions are poor.

(3) Real-time

The real-time nature of the translation system is a very important performance. To ensure the real-time nature of the system, it is necessary to ensure the response speed of the system and the speed of the translation process. In order to ensure the system response speed, the client memory and client design need to be optimized. At the same time, the translation algorithm and translation interface design need to be optimized to achieve faster translation speed.

(4) Ease of use

Ease of use is the common goal of translation systems. In order for users to quickly and instantly access the translation system, the front desk needs a concise and clear design so that they can find the functions they need in the shortest possible time. It is also important for users to easily access the translation page.

(5) Aesthetics

The beauty of the user interface is also an important feature of today's systems. Its simple and intuitive page usually allows users to determine if they are using the product. Beautiful and clear design can improve users' awareness of the system, and will attract users' attention when using it. By controlling the page layout and image style of the page, it is possible to meet the aesthetic needs of users while satisfying the performance of the software.

3 Experiment

3.1 Web Layer

The web layer contains a user interface module. This part mainly includes user input interface, translation result display interface, and clear, modify and restore buttons. The entire interface should be simple and easy to use. This section adds functions such as clearing, modifying, and copying, mainly to improve user experience. The delete box in the input box is mainly to delete the text in the input box. Users will inevitably encounter input errors during use, so if you delete the key at this time, it may reduce unnecessary delete functions. You can use the module in the translation result display box to change the translation result. In many cases, automatic translation is more or less problematic. Adding this part can make manual changes based on automatic translation, which is the same as using machine translation as an auxiliary tool to reduce manual workload. After

manual correction, a satisfactory translation is finally obtained. The main function of the copy unit of the translation result display box is to copy the content of the translation result display box, because the translated translation needs to be copied and pasted into the document for editing.

3.2 Translation Service Layer

The translation service layer includes the task scheduling part, the preprocessing part of the input sentence to be translated, the translation model part and so on. A task scheduler is added to coordinate the translation requests of multiple users to obtain better translation service resources and a better customer experience. Since this is a one-way translation from English to Chinese, the preprocessing of the data here only includes the preprocessing of the received English sentences to be translated, and the data is the input required by the translation model. The translation model here is an English translation model trained at the model service level. After the sentence to be translated is translated by the translation model, the translation result is fed back to the WEB layer, and the translation result is displayed to the client.

3.3 Model Training Layer

The training layer is mainly to complete the training task of the model, which is only visible to the server operator. Semi-supervised learning is selected in the training process. In the future, if the corpus on the server is expanded, only the translation model should be used as the pre-training model, and then the latest translation model can be obtained by training according to the algorithm. After the training is completed, the training layer delivers the pre-processing tools and copies of the pre-training results to the translation service layer along with the trained model.

3.4 Data Layer

The data preprocessing part is divided into translation service level preprocessing part and training level preprocessing part. Since the system is an English translation system, the translation requests processed are only for English translation requests. Therefore, the preprocessing part of the translation service level only deals with English sentences. The training-level preprocessing part includes English preprocessing part and Chinese preprocessing part. The English preprocessing modules of the two levels are similar. The configuration file of the English preprocessing part of the translation service is taken from the English preprocessing part of the training level, and is saved in the translation service layer as a copy of the result of the preprocessing part of the training level. Used for faster preprocessing tasks. The main purpose is to store configuration file information such as preprocessing vocabulary of the training layer. It is used to translate the service-level fast preprocessing feature to avoid server response delay caused by preprocessing and performance. The training-level data preprocessing part uses open source tools for word segmentation, BPE encoding, subtraction vocabulary extraction, word vector training and other functions, and the information is stored in the configuration file.

4 Discussion

In order to verify the effectiveness of the designed system, the performance of the translation system of this system and the other two systems are compared. The experimental data in this paper uses a certain machine translation task data set, which includes training data about 10 million Chinese-English parallel sentence pairs, and 8000 sentences each in the development set and the test set. During the experiment, 100,000 parallel sentences in the training set were randomly selected as training data, and the length of the source language and target language sentences was restricted to no more than 40 words.

Table 1. Comparison of translation performance of different sentence lengths

	0-10	10-20	20-30	30-40	40-50
Text system	24.51	25.3	25.56	25.62	24.71
System 1	23.85	23.9	16.3	14.26	11.35
System 2	24.4	24.54	17.41	15.44	13.04

The comparison results of the translation performance of each translation system for different sentence length distributions are shown in Table 1. In order to more intuitively show the changes in the performance of the translation model at different sentence lengths, the experimental results are shown in the form of a histogram as shown in Fig. 2.

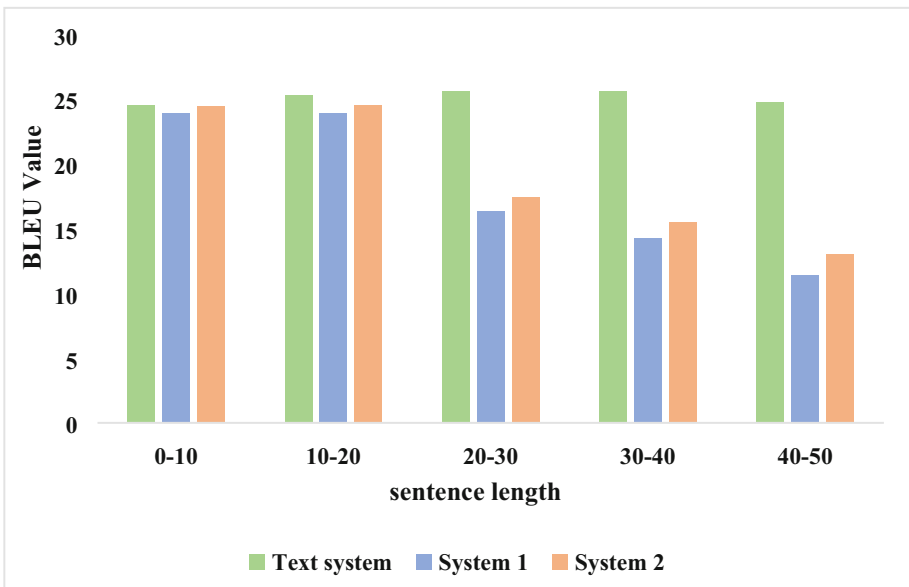


Fig. 2. Comparison of translation performance of different sentence lengths

By analyzing Fig. 2, it can be found that no matter in any length interval, the performance of the translation system in this paper is optimal, and as the sentence length gradually increases, the translation performance gradually increases.

5 Conclusions

Natural language processing is one of the most important topics in artificial intelligence technology research, and it is very difficult to implement in actual work. Machine translation is one of the biggest difficulties in processing natural language, and many difficulties have also been encountered in the research process. However, with the development of neural network technology, with the help of this technology, the effect of English machine translation has been significantly improved, which undoubtedly opens a new window for the research of English machine translation.

Acknowledgements. Research Project Supported by the National First-class Major Construction of Xi'an Fanyi University(Grant No. 21WYZA04).

References

1. Zhang, H.: Neural network-based tree translation for knowledge base construction. *IEEE Access*, **99**, 1–1 (2021)
2. Ren, Q., Su, Y., Wu, N.: Research on Mongolian-Chinese machine translation based on the end-to-end neural network. *Int. J. Wavelets Multiresolut. Inf. Process.* **18**(01), 46–59 (2020)
3. Maoxi, L.L., Xiang, Q., Chen, Z., et al.: A unified neural network for quality estimation of machine translation. *IEICE Trans. Inf. Syst.* E101.D(9), 2417–2421 (2018)
4. Xu, J., Wang, Z., Chao, T., et al.: Cutting pattern identification for coal mining shearer through a swarm intelligence-based variable translation wavelet neural network. *Sensors* **18**(2), 382 (2018)
5. Alanis, A.Y.: Electricity prices forecasting using artificial neural networks. *IEEE Latin Am. Trans.* **16**(1), 105–111 (2018)
6. Kahng, M., Andrews, P.Y., Kalro, A., et al.: ActiVis: visual exploration of industry-scale deep neural network models. *IEEE Trans. Vis. Comput. Graph.* 1 (2018)
7. Dong, Y., Wang, H.: Robust output feedback stabilization for uncertain discrete-time stochastic neural networks with time-varying delay. *Neural Process. Lett.* **51**(1), 83–103 (2019). <https://doi.org/10.1007/s11063-019-10077-x>
8. Anible, B.: Iconicity in American Sign Language-English translation recognition. *Lang. Cogn.* **12**(1), 138–163 (2020)
9. Zhang, X.: A study of cultural context in Chinese-English translation. *Region – Educ. Res. Rev.* **3**(2), 11–14 (2021)
10. Liu, Y.: Automatic proofreading method for English translation accuracy of nano vocabulary. *Nanotechnol. Environ. Eng.* **6**(3), 1–7 (2021). <https://doi.org/10.1007/s41204-021-00145-w>
11. Lucas, A., Iliadis, M., Molina, R., et al.: Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Process. Mag.* **35**(1), 20–36 (2018)
12. Xie, T., Grossman, J.C.: Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**(14), 145301.1–145301.6 (2018)



Design of Glasses Products Based on Artificial Intelligence

Han Sun and KieSu Kim^(✉)

School of Design, Silla University, Busan 46958, Korea

Abstract. With the rapid development of artificial intelligence and information technology, the way people obtain and process information has changed from a single way to a differentiated and personalized multiple way. As a means of transmitting and exchanging man-machine information, smart glasses have developed rapidly in recent years and have been widely reported by the media. The purpose of this article is to study the design of glasses products based on artificial intelligence. This article summarizes the functional requirements of smart glasses products through the questionnaire survey method, and explains the relevant module design of the products. Then carry out the design practice and usability evaluation of the product, by 3D printing the smart glasses product, carrying out the usability test on the printed white mold, and adjusting the design plan accordingly through the test results. According to survey data, 82.44%, 69.89%, 46.95%, and 59.86% of the number of people who believe that products should have the characteristics of durability, strong equipment, face recognition, and voice recognition account for 82.44%, 69.89%, 46.95%, and 59.86%. Therefore, durable, strong equipment, voice recognition, the feature of firm wearing are the main design requirements in the design practice of smart glasses products, and these features or functions need to be met first in the design.

Keywords: Augmented reality · Smart glasses · Industrial design · Appearance elements · Face recognition

1 Introduction

Smart glasses are a kind of mobile devices that have become popular at home and abroad in recent years. They can be divided into many types in terms of appearance, structure and function [1, 2]. There are single-center smart glasses, smart glasses, and holographic smart glasses that use near-eye technology, and some are just tools for collecting information [3, 4]. At present, the research of smart glasses has involved tourism, medical treatment, outdoor sports, police equipment, education and other application fields, which is of great significance to the research of smart glasses products [5, 6].

Regarding the research of smart glasses, many scholars at home and abroad have conducted multi-faceted discussions on them. For example, Jarosz J triggers from virtual reality to study the design of the interactive interface of glasses, which provides a basis for the interactive design of smart glasses [7]; Mitchell JP put forward some problems and key technology research on eye movement interaction [8]; Loh D explored

the development trend of smart glasses in appearance design, analyzed the influencing factors affecting appearance design elements, and proposed corresponding design principles [9].

This article focuses on the design of glasses products based on artificial intelligence. This article first analyzes the functional requirements of smart glasses products through the questionnaire survey method, and explains the relevant module design of the product, including. Finally, the design practice and usability evaluation of the product are carried out. Through the 3D printing of smart glasses products, the usability test of the printed white mold is carried out, and the design plan is adjusted accordingly through the test results.

2 Design of Glasses Products Based on Artificial Intelligence

2.1 Demand Analysis of Glasses Products Based on Artificial Intelligence

Image data collection: collect image data through the camera module;

Data storage: Temporarily store the data collected by the camera;

Wireless transmission: use Bluetooth/WIFI to transmit data;

Face recognition: After the system issues an instruction, it can perform face detection and recognition on the collected data;

Voice interactive control: users can control smart glasses through voice;

Photographing and video recording: the basic functions of smart glasses are realized through the camera module.

2.2 Related Module Design

(1) Microphone

The microphone device is implemented with a digital silicon microphone model AM4311T42A0. In this circuit design, two silicon microphones are used to fully achieve the noise reduction effect. One silicon microphone is connected to the main board, and the other silicon microphone is connected to the flexible. In addition, on the FPC, in this circuit, the silicon microphone is designed to be omnidirectional, that is, no directivity, and noise reduction can only be achieved through the cooperation of two silicon microphones. The sensitivity of this silicon microphone is -42 dB, and the signal-to-noise ratio is 58 dB. The comparison shows that the sensitivity and signal-to-noise ratio of this type of silicon microphone meet the design specifications. The noise figure is 36 dB (A)SPL, the working voltage is in the range of 1.5–3.6 V, and the current is 130 μ A when the 2 V voltage is in normal operation.

(2) Camera

Choose OV13850 CMOS sensor, this CMOS has low cost, sufficient information and convenient development. OV13850 collects the image into the M200 main control chip and stores it in the data buffer circuit, then transmits the image to the mobile phone through the M200 control module, and then displays it on the mobile phone.

(3) Camera interface circuit design

The smart glasses of this design use the OV13850 camera, which can programmatically control the image gain, exposure, frame rate, image size, horizontal mapping, vertical flipping and cropping, and translation; OV13850 supports image quality control and can correct image defects. Autofocus, the OV13850 camera itself supports MIPI architecture and can be connected to the MIPI camera interface provided by the main control chip.

(4) Image data cache

The product mainly designed two avatar data cache modules, one is the 8G cache that comes with the main control chip, but the cache in the system does not meet the data transmission requirements of the entire system, so the OV13850 chip is added, the camera circuit is designed, and the data cache. In the hardware circuit design of smart glasses, the NANDFlash flash memory interface supported by Ingenic M200 is used for 16-bit data transmission.

(5) LCD interface touch screen

The module's touch screen is placed inside the smart glasses shell, and a row of through holes is designed on the smart glasses shell. The touch screen is placed inside the through holes. The through hole design can meet the full contact of the touch screen with the outside world and increase the user's fingers [10, 11]. The interactive mode of the touch screen control is mainly used to control the functions of the smart glasses such as taking pictures, video recording, and voice calls, and is mainly controlled by the user's touch mode.

(6) WIFI and Bluetooth

The WIFI/Bluetooth module can be Bluetooth transmission, WIFI transmission, or both simultaneous transmission [12]. The WIFI/Bluetooth wireless transmission module mainly realizes data transmission between smart glasses and mobile phones and other terminals, and can also realize the binding between smart glasses and mobile terminals such as mobile phones, so that smart glasses can be connected to one or more devices.

WIFI positioning technology is a positioning technology based on receiving WIFI/signal strength, that is, the position of WIFI/hotspot is fixed, and the distance between the two is estimated according to the strength of the WIFI/signal received by the measured object, and then according to different WIFI and measured the distance of the object calculates the position of the measured object.

Bluetooth positioning technology is suitable for indoor positioning in a small area, but in a complex space environment, the stability of the system is poor and it is greatly affected by environmental noise.

(7) Wireless communication module

The current mainstream wireless transmission methods include Bluetooth, WIFI, FM and NFC, and the Newton platform integrates these four methods into one module, called WM-BAN-BM-13. Bluetooth supports 4.0 protocol, and also supports BER&ALE, low power consumption mode; smart glasses do not use Bluetooth 4.0, so this system still uses Bluetooth 3.0 to transmit data.

(8) Interactive mode of speech recognition:

In the voice input mode, the smart glasses system converts the vocabulary content in the user's voice into a computer-readable text string, which is convenient for the system to calculate and recognize. Therefore, the noise level of the user interface is one of the important factors affecting voice input. In a low-noise environment, the recognition of the voice input method is higher; in a noisy environment, the effect of voice input is easily affected by environmental factors. Therefore, generally speaking, smaller voice input commands have more accurate recognition than continuous voice input commands.

- 1) Execute simple commands: The basic functions of voice interaction in this smart glasses include the execution of some simple commands, such as opening an application and quick search.
- 2) Recognition and recording: This smart glasses design can help users recognize the correct semantic information of users, and help users to record voice content in various forms.
- 3) Identify identity: due to the difference in the user's voiceprint information, the voice interaction between smart glasses and the user can adopt voiceprint recognition technology. This technology can quickly identify the user's voice characteristics, help users quickly log in, but also prevent others from pretending to log in.

(9) Shell structure design

- 1) Considering that users who wear myopia glasses can wear them comfortably, a sponge cushion is set on the edge of the shell, so that the smart glasses can be worn comfortably while wearing myopia glasses. This is a very user-friendly design.
- 2) The two temples are connected with the front end of the shell by movable screw connection assembly, which makes the distance between the two temples adjustable. This design makes the smart glasses suitable for people with different sizes of faces.

3 Experimental Research on the Design of Glasses Products Based on Artificial Intelligence

3.1 Questionnaire Survey

User research first uses questionnaires to understand specific groups of people, quantify the daily work conditions of target users, and make the next stage of user interview questions more focused. In response to this research, the questionnaire “Questionnaire on the Demand for Smart Glasses Products Based on Artificial Intelligence” was designed. This questionnaire was distributed in both online and offline methods. A total of 300 questionnaires were distributed, including 80 on-site questionnaires and 120 online questionnaires. A total of 279 valid questionnaires were received.

3.2 T-Test

This article uses SPSS22.0 software to count and analyze the survey results, and perform t test. The t-test formula used in this article is as follows:

$$t = \frac{\bar{X} - \mu}{\frac{\sigma_X}{\sqrt{n}}} \quad (1)$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (2)$$

3.3 Product Testing

The content of this usability test includes six elements: ease of operation, wearing comfort, shape, color, function, and acceptance.

The main evaluation methods used this time are user model method and user review method. The user model method is mainly to 3D print smart glasses into a white model of the product, so that users can actually experience wearing them, simulate the user’s use scene and use process, observe the user’s behavior during the use of the product, and allow users to operate the product conveniently and wear the product. The user review is mainly to allow the target users to score the product shape, product color, and user acceptance.

Select the 6 elements of products that need user evaluation, and make a user evaluation questionnaire. The evaluation score is increased from 1 to 5 at a time. The higher the score, the higher the satisfaction. Users score based on their satisfaction with the product, and optimize the design through user evaluation results.

4 Data Analysis of Glasses Products Based on Artificial Intelligence

4.1 User Needs

Before designing a product, it is necessary to transform user requirements into design requirements to guide design practice. Counting the results of the questionnaire for user

needs, we get Table 1. The number of people who believe that the product has the characteristics of durability, strong equipment, face recognition, and voice recognition accounted for 82.44%, 69.89%, 46.95%, and 59.86%, respectively.

Table 1. Product performance requirements results

Element	Number of people	Proportion (%)
light	120	43.01
Strong equipment adaptability	195	69.89
Wear firmly	131	46.97
Precise positioning	117	41.94
Face recognition	131	46.95
Obstacle detection	124	44.44
Speech Recognition	167	59.86
durable	230	82.44
Photograph	100	35.84

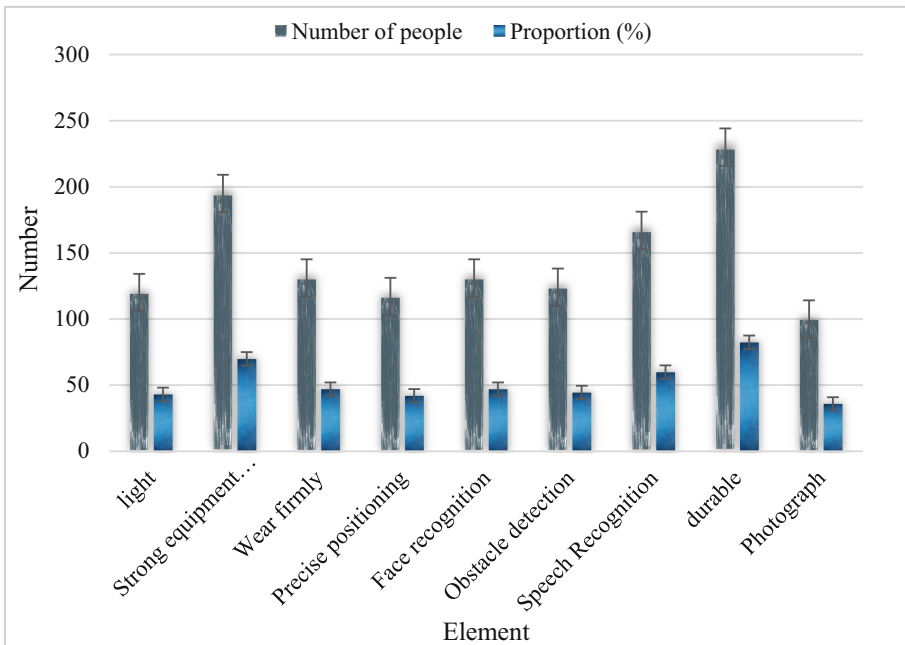


Fig. 1. Product performance requirements results

According to Fig. 1, it can be seen that the characteristics of durability, strong equipment, voice recognition, and strong wearing are the main design requirements in the

design practice of smart glasses products, and these characteristics or functions need to be met first during the design.

4.2 Product Availability

According to the user evaluation questionnaire scoring and the user’s oral presentation of questions after wearing, the results are shown in Table 2: 82, 78, 112, and 117 people scored 5 points for the product’s ease of operation, wearing comfort, shape, and color performance.

Table 2. User evaluation score table

Element	1	2	3	4	5
Ease of operation	0	17	56	124	82
Wearing comfort	0	56	72	73	78
modeling	0	17	75	75	112
color	20	54	33	75	117
Function	20	17	75	60	107
Acceptance	14	53	112	26	74

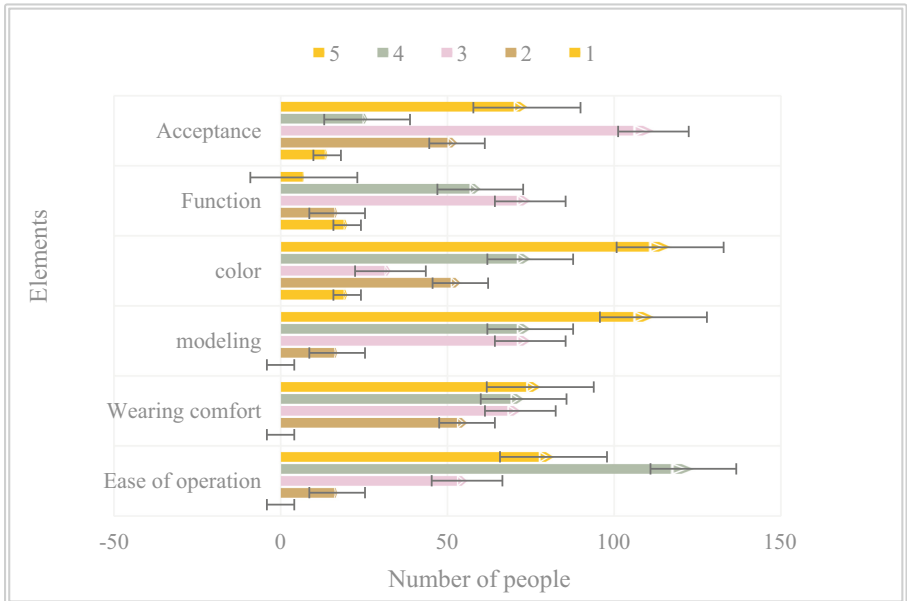


Fig. 2. User evaluation score table

It can be seen from Fig. 2 that users are more satisfied with the shape and color of the product because users are more inclined to choose bright product colors, which

helps relieve visual fatigue. At the same time, users have a low score on the wearing comfort of the product, indicating that the wearing characteristics of the product need to be improved.

5 Conclusion

With the development of technology, smart glasses products have entered a new stage as a new product design. This paper takes smart glasses product design as a research topic, analyzes the user needs of smart glasses development and design, and gives a design plan for smart glasses products. Finally, through a questionnaire survey, the product improvement plan is given. It is hoped that this research can help users who are ready to invest in the development of smart glasses, or who are committed to improving the production and service process, and play a good reference role.

References

1. Rauschnabel, P.A., Ro, Y.K.: Augmented reality smart glasses: an investigation of technology acceptance drivers. *Int. J. Technol. Mark.* **11**(2), 123–148 (2016)
2. Defaria, M., Oliveiradantas, D., Ruizalves, M., et al.: Visual acuity screening, photoscreening and dispensing of glasses with ready to clip technology. *Rev. Bras. Oftalmol.* **79**(5), 309–314 (2020)
3. Herz, M., Rauschnabel, P.A.: Understanding the diffusion of virtual reality glasses: the role of media, fashion and technology. *Technol. Forecast. Soc. Change* **138**, 228–242 (2019)
4. Kaylor, J., Hooper, V., Wilson, A., et al.: Reliability testing of augmented reality glasses technology: establishing the evidence base for telewound care. *J. WOCN* **46**(6), 1 (2019)
5. Klein, L., Aparicio, M., Jitianu, A.: *Handbook of Sol-Gel Science and Technology II Thermal Analysis on Gels, Glasses, and Powders*, pp. 1833–1867 (2018). https://doi.org/10.1007/978-3-319-32101-1_99
6. Marchi, J.: [Advanced Structured Materials] Biocompatible Glasses Volume 53 II Biocompatible Glasses for Controlled Release Technology, pp. 285–315 (2016). https://doi.org/10.1007/978-3-319-44249-5_12
7. Jarosz, J., MollieX, N., Chenon, G., et al.: Adaptive eyeglasses for presbyopia correction: an original variable-focus technology. *Opt. Express* **27**(8), 10533–10552 (2019)
8. Mitchell, J.P., Saraco, A.N., Ball, A.K., et al.: Breaking down barriers to learning with stereoscopic 3D technology: comparison of google cardboard, red-cyan anaglyphs & glasses-free 3D. *FASEB J.* **34**(S1), 1 (2020)
9. Loh, D., Zihajehzadeh, S., Hoskinson, R., et al.: Pedestrian dead reckoning with smartglasses and smartwatch. *IEEE Sens. J.* **PP**(22), 1 (2016)
10. Hirth, M.: gPhysics—using smart glasses for head-centered, context-aware learning in physics experiments. *IEEE Trans. Learn. Technol.* **9**(4), 304–317 (2016)
11. Iversen, M.D., Kiami, S., Singh, K., et al.: Prospective, randomised controlled trial to evaluate the effect of smart glasses on vestibular examination skills. *BMJ Innov.* **2**(2), bmjinnov-2015-000094 (2016)
12. Spitzer, M., Nanic, I., Ebner, M.: Distance learning and assistance using smart glasses. *Educ. Sci.* **8**(1) (2018)

Author Index

A

Almajed, Rasha, 3, 448
Altrjman, Chadi, 879, 971
Alturjman, Sinem, 45, 175

B

Bai, Jinniu, 1035
Bao, Changyuan, 787

C

Cai, Luyuan, 311
Cai, Xiao, 471
Cao, Shuai, 311
Cao, Yang, 537
Cao, Yimei, 754
Chen, Guilan, 384
Chen, Long, 78
Chen, Yanqing, 248
Cheng, Nan, 28, 400
Cheng, Xiaowei, 112
Chi, Xueqian, 853
Comstock, Allan, 669
Cui, Jianzhou, 693
Cui, Yu, 987

D

Dai, Liyun, 392
Di, Zilong, 480
Diao, Wei, 739
Ding, Yang, 569
Dong, Chunyang, 787
Dong, Jinfeng, 601
Dong, Qinwei, 166

F

Feng, Youjun, 216
Fu, Yi, 128

G

Gao, Jiawen, 471
Gao, Jing, 626
Ge, Lianying, 143
Gu, Yao, 151
Guo, Hui-yong, 278
Guo, Meina, 778
Guo, Xiaolan, 408

H

Hang, Feilu, 270
Hao, Lihong, 809
He, Ying, 715
He, Zijuan, 1027
Hou, Kenan, 787
Hou, Shuwei, 787
Hou, Yuqing, 1035
Hu, Yingyue, 45
Huang, Cuicui, 684
Huang, Jing, 311
Huang, Yi, 256
Hui, Fang, 368

J

Jain, Amar, 295, 978
Jain, Neha, 408, 569
Ji, Xiaopeng, 971
Jiang, Chengling, 166
Jiang, Guoqing, 601
Jiang, Nan, 303

Jiang, Zerong, 954
 Jiao, Jing, 846
 Jiejue, Yihong, 318

K

Kim, KieSu, 1051
 Kong, Yuqiang, 862
 Kuang, Zijia, 240

L

Lavoie, Thomas, 651
 Li, Depeng, 448
 Li, Hongyan, 52
 Li, Huajun, 37
 Li, Jiangnan, 771
 Li, Mingfu, 28
 Li, Na, 930
 Li, Qian, 192
 Li, Ruijuan, 723
 Li, Xia, 432
 Li, Xiaoliang, 327
 Li, Xiaoxia, 216
 Li, Ying, 684
 Li, Yu, 802
 Li, Zichen, 651
 Lin, Jinbang, 471
 Lin, Jing, 233
 Lin, Xianjin, 826
 Ling, Ruiliang, 978
 Liu, Chao, 112
 Liu, Chunling, 871
 Liu, Cui, 617
 Liu, Dapeng, 684, 731
 Liu, Hua, 601
 Liu, Jianjun, 787
 Liu, Juan, 96
 Liu, Le, 641
 Liu, Lin, 135
 Liu, Ling, 609
 Liu, Ming, 601
 Liu, Piao, 216
 Liu, Tao, 809
 Liu, Tengjiao, 497
 Liu, Wei, 448
 Liu, Weimeng, 11
 Liu, Wenbin, 256
 Liu, Zenghui, 416
 Lou, Jinman, 216
 Lu, Xiao, 158
 Lu, Yue, 505
 Luo, Xiu, 295
 Lv, Shasha, 61

M

Ma, Bing, 787
 Ma, Hongling, 424
 Ma, Jianwei, 175
 Ma, Le, 112
 Ma, Teng, 731
 Ma, Zhoujun, 166
 Mao, Chao, 837
 Metawa, Saad, 128, 318
 Ming, Dayang, 376

N

Niu, Zhiyuan, 562

P

Pan, Lei, 471
 Peng, Peilan, 455
 Peng, Yun, 270

Q

Qi, Yongjun, 651, 879
 Qian, Cheng, 343
 Qian, Shenghua, 488
 Qiu, Guihua, 240
 Qiu, Shi, 462

R

Ran, Zhuying, 922
 Ren, Beibei, 1019

S

Salave, Arnold, 676
 Sang, Hongyan, 660
 Shen, Xifeng, 601
 Shen, Yachen, 143
 Shi, Hongyun, 462
 Shi, Wei, 795
 Shukla, Shalini Kumar, 86, 497
 Su, Jizhi, 28, 400
 Sun, Han, 1051
 Sun, Xia, 896
 Sun, Xiangju, 112
 Sun, Yan, 19

T

Tan, Jing, 151
 Tang, Wei, 787
 Tang, Yingxia, 70
 Tang, Zhirui, 240
 Tao, Yu, 905
 Tian, Zhiguo, 739
 Tiwari, Rajan Singh, 303, 462
 Tu, Yihan, 1003

V

Vijayakumar, K., 143

W

Wan, Fangtao, 225
Wang, Chenhe, 676
Wang, Cong, 199
Wang, Dawei, 787
Wang, Gang, 537
Wang, Jun, 263
Wang, Quannan, 995
Wang, Shichun, 462
Wang, Tingwei, 199
Wang, Xiaoqing, 208
Wang, Xuyan, 938
Wang, Yamin, 400
Wang, Yan, 731
Wang, Yuanjie, 103
Wang, Yuqi, 184
Wang, Zhe, 158
Wei, Lei, 158
Wei, Taowen, 1011
Wu, Hailiang, 28
Wu, Haiyan, 879
Wu, Huihuang, 530

X

Xiang, Shishi, 562
Xiao, Chaoying, 754
Xiao, Longyuan, 593
Xiao, Xin, 112
Xie, Shengxiang, 225
Xie, Yuxin, 763
Xie, Zhengfeng, 462
Xing, Lin, 28
Xiong, Rong, 651
Xu, Li, 1035
Xu, Lianggang, 462
Xu, Xiaofeng, 787
Xue, Jia, 335

Y

Yan, Qiongqiong, 684
Yan, Weicheng, 392
Yang, Hongwei, 400
Yang, Jiali, 199
Yang, Jing, 746
Yang, Lielong, 700
Yang, Long, 462
Yang, Pingping, 351
Yang, Qiong, 585
Yang, Yaokun, 669

Yao, Zhuojun, 175
Ye, Jing, 86
Yi, Danping, 995
Yin, Xingtian, 700
Yu, Chenglong, 440
Yu, Chunjiang, 128
Yu, Jiang, 577
Yu, Kai, 216
Yu, Yanjie, 513
Yuan, Jianshu, 225

Z

Zang, Shilong, 471
Zhang, Congxiao, 913
Zhang, Dan, 311
Zhang, Hanxu, 963
Zhang, Lipeng, 634
Zhang, Liping, 327
Zhang, Mengxue, 3
Zhang, Qiong, 278
Zhang, Ronggen, 554
Zhang, Shaocheng, 199
Zhang, Shunji, 723
Zhang, Tao, 946
Zhang, Tianhao, 471
Zhang, Wuchen, 400
Zhang, Xinyi, 135
Zhang, Xu, 112
Zhang, Yan, 359
Zhang, Yuhao, 285
Zhao, Bo, 787
Zhao, Di, 754
Zhao, Fengjun, 1035
Zhao, Xin, 601
Zhao, Ying, 545
Zhao, Zheng, 1043
Zheng, Hui, 684, 731
Zheng, Lixia, 225
Zheng, Wang, 96
Zheng, Wei, 809
Zheng, Ziyu, 826
Zhong, Qiong, 817
Zhou, Bei, 521
Zhou, Cong, 887
Zhou, Jie, 505
Zhou, Jing, 731
Zhou, Jun, 537
Zhou, Lu, 987
Zhou, Zhuolin, 708
Zhu, Chen, 954
Zhu, Hui, 120