

Shuihua Wang  
Zheng Zhang  
Yuan Xu (Eds.)



414

LNICST

# IoT and Big Data Technologies for Health Care

Second EAI International Conference, IoTCare 2021  
Virtual Event, October 18–19, 2021  
Proceedings, Part I

Part 1



# Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering

414

## Editorial Board Members

Ozgur Akan

*Middle East Technical University, Ankara, Turkey*

Paolo Bellavista

*University of Bologna, Bologna, Italy*

Jiannong Cao

*Hong Kong Polytechnic University, Hong Kong, China*

Geoffrey Coulson

*Lancaster University, Lancaster, UK*

Falko Dressler

*University of Erlangen, Erlangen, Germany*

Domenico Ferrari

*Università Cattolica Piacenza, Piacenza, Italy*

Mario Gerla

*UCLA, Los Angeles, USA*

Hisashi Kobayashi


*Princeton University, Princeton, USA*

Sergio Palazzo

*University of Catania, Catania, Italy*

Sartaj Sahni

*University of Florida, Gainesville, USA*

Xuemin Shen 

*University of Waterloo, Waterloo, ON, Canada*

Mircea Stan

*University of Virginia, Charlottesville, USA*

Xiaohua Jia

*City University of Hong Kong, Kowloon, Hong Kong*

Albert Y. Zomaya

*University of Sydney, Sydney, Australia*

More information about this series at <https://link.springer.com/bookseries/8197>

Shuihua Wang · Zheng Zhang · Yuan Xu (Eds.)

# IoT and Big Data Technologies for Health Care

Second EAI International Conference, IoTCare 2021  
Virtual Event, October 18–19, 2021  
Proceedings, Part I

*Editors*

Shuihua Wang   
University of Leicester  
Leicester, UK

Zheng Zhang  
Harbin Institute of Technology  
Shenzhen, China

Yuan Xu   
University of Jinan  
Jinan, China

ISSN 1867-8211                      ISSN 1867-822X (electronic)  
Lecture Notes of the Institute for Computer Sciences, Social Informatics  
and Telecommunications Engineering  
ISBN 978-3-030-94184-0              ISBN 978-3-030-94185-7 (eBook)  
<https://doi.org/10.1007/978-3-030-94185-7>

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2022

This work is subject to copyright. All rights are reserved 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

# Preface

We are delighted to introduce the proceedings of the second edition of the European Alliance for Innovation (EAI) International Conference on IoT and Big Data Technologies for HealthCare (IoTCare 2021). This conference brought together researchers, developers, and practitioners around the world who are leveraging and developing technology for Internet of Things (IoT) and big data applications in healthcare. The theme of IoTCare 2021 was the convergence of IoT and big data technologies for e-health, e-care, lifestyle, aging populations, smart personal living applications, etc.

The technical program of IoTCare 2021 consisted of 80 full papers in the oral presentation sessions at the main conference track on integrating healthcare with IoT. Aside from the high-quality technical paper presentations, the technical program also featured two keynote speeches and two technical workshops. The two keynote speeches were given by Manu Malek from Stevens Institute of Technology, USA, and Ivan Tyukin, Turing Artificial Intelligence Fellow, from the University of Leicester, UK. The two workshops organized were AI-based Internet of Medical Things (IoMT) and Information Fusion for the Devices of Internet of Things (InfusIoT). IoMT aimed to address the combination of IoT and AI to enable more personalized, preventative, and collaborative forms of IoT care. InfusIoT aimed to address the issue of information fusion for IoT devices to maintain high accuracy, adaptiveness, robustness, timeliness, reliability, and intelligence.

Coordination with the steering chairs, Imrich Chlamtac and Liangxiu Han was essential for the success of the conference. We sincerely appreciate their constant support and guidance. It was also a great pleasure to work with such an excellent organizing committee team for their hard work in organizing and supporting the conference. In particular, we are grateful to the Technical Program Committee, who have completed the peer-review process for the technical papers and helped to put together a high-quality technical program. We are also grateful to Conference Manager Natasha Onofrei for her support and all the authors who submitted their papers to the IoTCare 2021 conference and workshops.

We strongly believe that the IoTCare conference provides a good forum for all researchers, developers, and practitioners to discuss all science and technology aspects that are relevant to IoT in healthcare. We also expect that the future IoTCare conferences will be as successful and stimulating as this year's, as indicated by the contributions presented in this volume.

November 2021

Shuihua Wang  
Yu-Dong Zhang

# Organization

## Steering Committee

Imrich Chlamtac  
Liangxiu Han

University of Trento, Italy  
Manchester Metropolitan University, UK

## Organizing Committee

### General Chair

Shui-Hua Wang

University of Leicester, UK

### General Co-chair

Shuai Liu

Hunan Normal University, China

### Technical Program Committee Chair

Yu-Dong Zhang

University of Leicester, UK

### Technical Program Committee Co-chairs

Jin Sun

Yangzhou University, China

Ruidan Su

Shanghai Advanced Research Institute,  
Chinese Academy of Sciences, China

Zheng Zhang

Harbin Institute of Technology, China

Juan Manuel Górriz Sáez

Universidad de Granada, Spain

Muhammad Attique Khan

HITEC University, Pakistan

### Sponsorship and Exhibit Chair

Zhanhan Tu

University of Leicester, UK

### Local Chair

Xiang Yu

University of Leicester, UK

### Workshops Chairs

Yuan Xu

University of Jinan, China

Zedong Zheng

De Montfort University, UK

### **Publications Chairs**

Chong Zeng University of Leicester, UK  
Xujing Yao University of Leicester, UK

### **Web Chair**

Siyuan Lu University of Leicester, UK

### **Technical Program Committee**

Chun Guang Henan Vocational College of Industry and  
Information Technology, China  
Chunli Guo Inner Mongolia University, China  
Cui Jianfeng Xiamen University of Technology, China  
Dan Zhang Xinyang Vocational and Technical College, China  
Dongye Liu Inner Mongolia University, China  
Feng Cheng Xizang Minzu University, China  
Gaocheng Liu Inner Mongolia University, China  
Hao Xu Xinyang Vocational and Technical College, China  
Huadong Wang Inner Mongolia University, China  
Jamal Hussain Shah COMSATS University Islamabad, Pakistan  
Jiangyi Zhang Jiangnan University, China  
Jianwei Zhang Zhengzhou University of Light Industry, China  
Junaid Ali Khan HITEC University, Pakistan  
Kashif Javed National University of Science and Technology,  
Pakistan  
Keming Mao Northeastern University, China  
Li Heng Henan Finance University, China  
M. Hassaballah South Valley University, Egypt  
Ma Lei Beijing Polytechnic, China  
Mingcheng Peng Jiangmen Vocational and Polytechnic College,  
China  
Mudassar Raza COMSATS University Islamabad, Pakistan  
Muhammad Attique Khan COMSATS University Islamabad, Pakistan  
Muhammad Sharif COMSATS University Islamabad, Pakistan  
Muhammad Younus Javed HITEC University, Pakistan  
Na Ta Inner Mongolia University, China  
Rajinikanth Venkatesan St Joseph's College of Engineering, India  
Rameez Naqvi COMSATS University Islamabad, Pakistan  
Reham Mostafa Mansoura University, Egypt  
Robertas Damasevicius Vytautas Magnus University, Lithuania  
Shuai Liu Hunan Normal University, China  
Shuai Yang Changchun University of Technology, China



Shuihua Wang	University of Leicester, UK
Siyuan Lu	University of Leicester, UK
Sui Dan	Cal Poly Pomona, USA
Tallha Akram	COMSATS University Islamabad, Pakistan
Tenghui He	Hunan Normal University, China
Tian Hong	Baotou Iron and Steel Vocational Technical College West Gate, China
Tong Xuanyue	Nanyang Institute of Technology, China
Vishnu Varthanan	Kalasalingam Academy of Research and Education, India
Weiling Bai	Inner Mongolia University, China
Wenda Xie	Jiangmen Polytechnic, China
Xiang Yu	University of Leicester, UK
Xiaogang Zhu	Nanchang University, China
Xinchun Zhou	Baoji University of Arts and Sciences, China
Xinyu Liu	Hunan Normal University, China
Xuechao Zhang	Hulunbuir Vocational Technical College, China
Xujing Yao	University of Leicester, UK
Yanning Zhang	Beijing Polytechnic, China
Yi Chen	Nanjing Normal University, China
Yuling Jin	Chizhou Vocational and Technical College, China
Zhengchao Dong	Columbia University, USA

# Contents – Part I

## Integrating Healthcare with IoT

Research on the Universal Access Security Authentication Technology of Multi-source Heterogeneous Terminal Communication Module .....	3
<i>Bao-ren Chen, Dan-ke Hong, Li Wang, Yong-tong Ou, and Xin-hui Zhong</i>	
Research on Intelligent Operation and Maintenance Technology of Pumped Storage Power Plant Based on 5G .....	15
<i>Meng Ye, Guan-jin Huang, Peng Gao, Miao-geng Wang, and Xu-hui Zhang</i>	
Smart Grid Service Transmission Accuracy Optimization Technology Based on 5G Technology .....	27
<i>Li-kuan Gong, Jian-yong Zhou, Guo-yi Zhang, Tong-hao Wu, and Zhi-wei Liu</i>	
Research on Hybrid Communication Networking Protocol Optimization Technology of Ubiquitous Electric Internet of Things .....	40
<i>Bao-ren Chen, Dan-ke Hong, Li Wang, Yong-tong Ou, and Xin-hui Zhong</i>	
Forecasting Method of Monthly Clearing Price Under the Background of Continuous Adjustment of Power Market Supply and Demand .....	53
<i>Guo-bin Wang, Wen-tao Xu, Le-le Wang, Jing An, and Yang Bai</i>	
Automatic Response Method of Grid Demand Under the Background of New Energy Consumption .....	71
<i>Wen-tao Xu, Chong-wei Ma, Guo-bin Wang, Ye-fei Li, and Hui Liu</i>	
Effect Evaluation of College English Blended Teaching Based on Factor Analysis and Association Rules .....	88
<i>Rong-rong Ruan and Meng Ye</i>	
Design of Golfer's Heart Rate Data Transmission System Based on Machine Learning .....	105
<i>Bin-bin Liu</i>	
Design of Healthcare Data Analysis System Based on Operational Research and Differential Evolution Algorithm .....	119
<i>Xue Jin and Bin-bin Liu</i>	

Research on Reliability Evaluation of Intelligent Energy Integrated Service Platform Based on Hierarchical Linear Model .....	136
<i>Wen-Lin Xu, Xiao-xiao Liu, Chao Li, Zi-peng Hu, and Kai Liu</i>	
Design of Clinical Medical Data Monitoring System Based on Artificial Intelligence and Big Data .....	154
<i>Tao Lei and Gui-xiu Xie</i>	
Construction of Pediatric Medication Data Security Cloud Storage Model Based on Internet of Things Technology .....	171
<i>Shu-hua Whang, Wen-shan Yao, Xian-ying Meng, Min-na Zheng, and Hua Fan</i>	
Mathematical Model of Data Partition Storage in Network Center Based on Blockchain .....	187
<i>Bing-bing Han, Zai-xing Su, and Hai-yun Han</i>	
Enterprise Cluster Intelligent Manufacturing Information Management System Based on Wireless Communication Technology .....	200
<i>Yang Su, Yu-hua Gai, and Qiu-ying Lv</i>	
3D-HEVC Deep Video Information Hiding and Secure Transmission Method .....	212
<i>Cai-xu Xu, Hui Guo, Cai-cun Cen, and Yong-ming Chen</i>	
Research on Difference Elimination Method Between Small Sample Databases Based on Feature Extraction .....	226
<i>Jin-hua Liu and Fu-lian Zhong</i>	
Research on Homomorphic Retrieval Method of Private Database Secrets in Multi-server Environment .....	244
<i>Fu-lian Zhong and Jin-hua Liu</i>	
Low-Frequency Noise Characteristic Extraction Method of Electronic Components Based on Data Mining .....	260
<i>Xiao-jing Qi</i>	
X-band Radar Detection Target Tracking Method Based on Internet of Things Sensing Technology .....	272
<i>Fengshuo Yan, Kui Xiong, Mingyang Gao, and Yanxin Shu</i>	
Research on Hybrid Encryption of Cross Border E-commerce Transaction Information Based on B+ Search Tree Algorithm .....	290
<i>Jia-hua Li</i>	

Research on Fuzzy Retrieval Method of Blockchain Information Based on B+Tree Index Structure .....	308
<i>Jia-hua Li</i>	
Reliability Analysis Method of Multi Area Fault Diagnosis and Location in Power Grid with Missing Information .....	326
<i>Jun-ci Tang, Tie Li, Jun-bo Pi, Miao Wang, and Feng Jiang</i>	
Reliability Analysis of Real Time Operation State of Power System Based on K-Nearest Neighbor .....	344
<i>Tie Li, Jun-ci Tang, Jian-ming Yu, Dai Cui, and Di Jiang</i>	
Low Resolution License Plate Recognition Based on Intelligent Data Processing and Prediction Algorithm .....	361
<i>Mi Meng and Chun-hu He</i>	
Design and Application of Security Monitoring System for Perception Terminal of Power Internet of Things .....	373
<i>Hong Xu, Xin Sun, Jie-yao Ying, and Qing-li Niu</i>	
Reliability Evaluation Model of Online Teaching Quality Based on Big Data Technology .....	386
<i>Feng Wang</i>	
Reliability Detection Method of Online Education Resource Sharing Based on Blockchain .....	399
<i>Feng Wang</i>	
Research on Inventory Control Method of Cold Chain Logistics Enterprises Under the Background of New Energy Consumption .....	417
<i>Dong-ming Yue, Hui-ling Zhang, Guang-yao Miao, Le-le Wang, and Feng-shuo Yan</i>	
Fault Detection Method of Electronic Circuit Converter Based on Dynamic Sequence Response .....	433
<i>Dan-kang He and Qiu-jiao Huang</i>	
Reliability Analysis of Power Side Information Acquisition Model Based on Wireless Sensor .....	447
<i>Wen-Lin Xu, Xin-Ze Guo, Zi-Peng Hu, Chao Li, and Kai Liu</i>	
Research on Obstacle Avoidance Tracking Planning of Hyper-redundant Manipulator Based on VR Technology .....	464
<i>Xiao-zheng Wan, Song Zhang, Ji-ming Zhang, Hui Chai, and Dong-bao Ma</i>	

High Voltage Cable Shield Voltage Monitoring Method Considering Coil Flux .....	476
<i>Qiu-jiao Huang and Dan-kang He</i>	
Design of Construction Cost Over Budget Control System Based on BIM Technology .....	492
<i>Yun-xia Xie and Miao Wei</i>	
Cluster Evaluation Method of Building Decoration Cost Rationality Based on BIM Model .....	509
<i>Miao Wei and Yun-xia Xie</i>	
Power Frequency Control Method of Electrical Equipment Under Internet of Things Technology .....	526
<i>Hong-yu Huang and Qing-huan Qin</i>	
PID Based Automotive Electronically Controlled Power Steering System in the Internet of Things Environment .....	542
<i>Qing-huan Qin and Hong-yu Huang</i>	
Research on Debounce Method of Electronic Imaging Equipment Based on Feature Point Matching .....	559
<i>Xiao-jing Qi</i>	
An Incomplete License Plate Image Intelligent Recognition System Based on the Generated Counter Network .....	570
<i>Mi Meng, Chun-hu He, and Xiao-jing Qi</i>	
<b>Author Index</b> .....	583

## Contents – Part II

### Integrating Healthcare with IoT

Data Transmission Reliability Detection of Hybrid Information System Based on Smart Contract .....	3
<i>Huan-yu Wang and Xiao-gang Ma</i>	
Research on Network Security Authentication Method Based on Data Mining Technology .....	19
<i>Xiao-gang Ma and Huan-yu Wang</i>	
High Reliability Design of Student Status Information Acquisition System in Ideological and Political Classroom Under Multi-target Tracking .....	30
<i>Gui-xiu Xie and Tao Lei</i>	
Weak Vibration Signal Extraction Method of Mechatronics Equipment Based on Stochastic Resonance .....	46
<i>Dong-bao Ma, Xue-mei Li, Ming-fei Qu, and Xiao-zheng Wan</i>	
Fast Integration System of English Online Learning Resources Based on Multi Sensor Network .....	58
<i>Hai-yun Han and Bing-bing Han</i>	
Quality Evaluation of Human Resource Management Information System Based on Intelligent Optimization Algorithm .....	71
<i>Bo Sun and Hao-nan Chu</i>	
Human Resource Social Insurance Data Dynamic Update System Based on Wireless Communication .....	88
<i>Hao-nan Chu and Bo Sun</i>	
Design of Wireless Audio Real Time Transmission Model Based on Body Area Network Technology .....	103
<i>Qing-li Niu and Hong Xu</i>	
Time-Frequency Analysis of Vibration Signal Distribution of Rotating Machinery Based on Machine Learning and EMD Decomposition .....	115
<i>Xiao-zheng Wan, Song Zhang, Ji-ming Zhang, Hui Chai, and Huan-yu Zhao</i>	

Synchronous Monitoring Method of Multi-manipulator Trajectory Signals Based on Machine Learning .....	129
<i>Xiao-zheng Wan, Song Zhang, Ji-ming Zhang, Hui Chai, and Huan-yu Zhao</i>	
Big Data Stepwise Regression Correction Method for Forearm Wrong Posture in Track and Field .....	140
<i>Yong-ming Chen and Cai-xu Xu</i>	
Intelligent Sharing Technology of Mobile Medical Dynamic Data Based on Internet of Things .....	153
<i>Hai-bo Zhang, Xiu-juan Duan, and Jian-mei Sun</i>	
Multi Channel Data Encryption Transmission Algorithm of Medical Internet of Things Based on Improved MQTT Protocol .....	171
<i>Hai-bo Zhang, Xiu-juan Duan, and Jian-mei Sun</i>	
Design of Innovation and Entrepreneurship Effect Evaluation System for College Students Based on MOA Model .....	183
<i>Dan Zhao and Huan-wei Liang</i>	
Design of Human Resource Distance Education System Based on Internet of Things Technology .....	199
<i>Huan-wei Liang and Dan Zhao</i>	
Intelligent Encrypted Storage Method for Medical Health Database Based on Internet of Things .....	216
<i>Qing-bang Zeng and Wen-da Xie</i>	
Design of Single Camera 3D Laser Scanning System Based on Artificial Intelligence .....	233
<i>Wen-da Xie and Qing-bang Zeng</i>	
Design of Enterprise Intelligent Decision Support System Based on Data Mining .....	249
<i>Qiu-ying Lv and Yang Su</i>	
Recommendation Method of Nursing Teaching Resources in Geriatric Internal Medicine Based on Internet of Things Technology .....	268
<i>Hua Fan</i>	
Recognition of Human Abnormal Behavior in Static Image of Intelligent Monitoring System Based on Neural Network Algorithm .....	280
<i>Hai-jing Zhou</i>	

Fall Behavior Recognition Algorithm in Video Surveillance Based on Feature and Deep Learning .....	298
<i>Hai-jing Zhou</i>	
Pneumonia Detection Algorithm Based on Improved YOLOv3 .....	313
<i>Hailong Liu, Jinrong Cui, and Chaoda Peng</i>	
Breast Ultrasound Images Clustering Analysis Using Deep Clustering Method .....	321
<i>Cheng Huang and Jinrong Cui</i>	
<b>Information Fusion for the Devices of Internet of Things</b>	
Comparing Methods of Imputation for Time Series Missing Values .....	333
<i>Renkang Geng, Mingran Li, Mingxu Sun, and Yujie Wang</i>	
Clustering-XGB Based Dynamic Time Series Prediction .....	341
<i>Haoxuan Sun, Kun Zhang, Tingting Wang, Wanfeng Ma, and Qinjun Zhao</i>	
Time Series Data Imputation Using Expectation-Maximization with Principal Component Analysis .....	352
<i>Renkang Geng, Jing Cao, Qinjun Zhao, and Yujie Wang</i>	
Time Series Prediction with Preprocessing and Clustering .....	358
<i>Haoxuan Sun, Shuai Lin, Lin Han, Jidong Feng, and Mingxu Sun</i>	
Using an Ensembled Boosted Model for IoT Time Series Regression .....	368
<i>Shuai Lin, Kun Zhang, Renkang Geng, and Liyao Ma</i>	
Dynamic Time Warping Based Clustering for Time Series Analysis .....	376
<i>Kun Zhang, Shuai Lin, Haoxuan Sun, Liyao Ma, and Junpeng Xu</i>	
Voltage Estimation of PH Meter Calibrator Using Integration of Kalman/FIR and R-T-S Smoothing Method .....	386
<i>Wanjie Ren, Xia Li, Guoxing Hu, Rui Tuo, and Chen Cai</i>	
Human Tracking Using Distributed Dual-EKF Filter .....	394
<i>Jing Cao, Jidong Feng, Wanjie Ren, Wanfeng Ma, Mingran Li, and Yuan Xu</i>	
Evaluation and Analysis of Basic-Level Aircraft Maintenance and Support Capabilities .....	402
<i>Yanli Gao, Shaokang Ji, Jianling Qu, and Mingran Li</i>	



<b>Robust Unscented Kalman Filter for Target Tracking Based on Mahalanobis Distance</b> .....	419
<i>Bingbing Gao, Wenmin Li, Longqiang Ni, and Wei Wang</i>	
<b>Analysis of Chaotic Behavior in Single Mode NH<sub>3</sub> Molecular Laser</b> .....	433
<i>Hongyan Zang, Shourong Zhang, and Tengfei Lei</i>	
<b>Collaborative Path Optimization Method for Flood Control Material Storage</b> .....	440
<i>Zhihao Li, Xinyao Wang, Shuhui Bi, and Qinjun Zhao</i>	
<b>Detection of Rail Bottom Damage Defects Based on Recurrent Neural Network</b> .....	451
<i>Fengguang Zhou, Qinjun Zhao, Yuan Xu, Qinhua Xu, and Tao Shen</i>	
<b>Short Term Load Forecasting Method Based on Full Convolution Deep Learning</b> .....	461
<i>Hai-hong Bian, Xing-jian Shi, Qian Wang, and Li-kuan Gong</i>	
<b>AI-Based Internet of Medical Things</b>	
<b>Covid-19 Detection by Wavelet Entropy and Cat Swarm Optimization</b> .....	479
<i>Wei Wang</i>	
<b>A Short Survey on Deep Learning Models for Covid-19 Detection Based on Chest CT and X-ray Images</b> .....	488
<i>Wei Wang</i>	
<b>IoT and AI Technology Used for COVID-19 Pandemic Control</b> .....	497
<i>Shu-Wen Chen and Xiao-Wei Gu</i>	
<b>Review of Covid-19 Diagnosis Techniques Combined with Machine Learning and AI Analysis</b> .....	508
<i>Xiao-Wei Gu, Shu-Wen Chen, Xuan Tong, Hui-Shen Yan, Lu Chen, and Si-Ye Wu</i>	
<b>Author Index</b> .....	523

# **Integrating Healthcare with IoT**



# Research on the Universal Access Security Authentication Technology of Multi-source Heterogeneous Terminal Communication Module

Bao-ren Chen<sup>1(✉)</sup>, Dan-ke Hong<sup>1</sup>, Li Wang<sup>1</sup>, Yong-tong Ou<sup>2</sup>,  
and Xin-hui Zhong<sup>2</sup>

<sup>1</sup> Communication Office of China Southern Power Grid,  
Guangzhou 510670, China  
cbr42561@163.com

<sup>2</sup> Digital Power Grid Branch of China Southern Power Grid Digital Power Grid  
Research Institute, Guangzhou 510670, China

**Abstract.** Access interference and security authentication of multi-source and heterogeneous terminals are always important factors that affect the successful communication of wireless networks. Because of the contradiction between the authentication efficiency of handoff and the access security, the security of the traditional multi-source heterogeneous terminal communication module cannot meet the application requirements. Aiming at this problem, this paper puts forward the research on the security authentication of multi-source heterogeneous terminal communication module. Using the uniform data communication protocol, we collect and process the data of multi-source heterogeneous terminal, and design the access security authentication protocol based on random number. Experimental results show that the authentication technology of multi-source heterogeneous terminal communication module has low authentication delay and good stability, and its overall security is improved.

**Keywords:** Multi-source heterogeneous terminal · Communication module · Pan-access · Security authentication technology

## 1 Introduction

With the rapid development of information science and technology and the popularization of network applications, especially the development of mobile networks, intelligent mobile terminals, mobile applications and other technologies, mobile office and remote operation tend to mature, making remote office more convenient and effective, and greatly improving the efficiency and efficiency of work [1]. At present, wireless network connection is not only a simple way of network access, but also a lot of important applications such as language service, video service and location-based service. Users pay more attention to the fluency of browsing after network connection, but with the rapid development of network multimedia application market, the number of various network applications increases rapidly, and the wireless interference

problems encountered by dynamic users in the process of network connection increase accordingly, which leads to the poor quality of network service and difficult to achieve the expected experience of users. In addition, due to the characteristics of mobile networks and terminals, information security problems caused by network viruses and hackers are becoming more and more important [2]. Especially in the field of healthcare settings authentication, network virus not only interferes with the communication process of medical information, but also causes the loss of patients' privacy information [3]. Therefore, it is urgent to study the secure access scheme based on mobile environment. The importance of information security is self-evident, especially for banks, State Grid, public security organs and other government agencies and national enterprises, information security is particularly important [4]. Therefore, how to ensure that confidential data will not be leaked, and to achieve the authentication of mobile access objects is also the most important security access scheme.

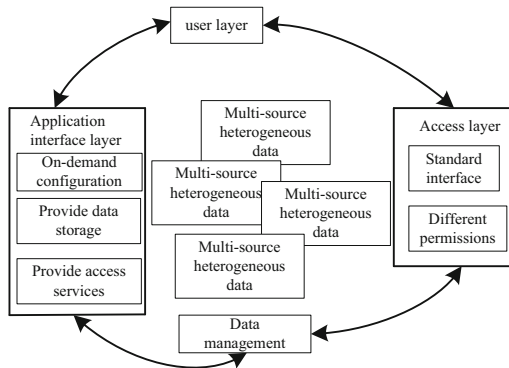
In reference [5], a security protection method of user login information authentication based on qos-afd was proposed. By analyzing the failure frequency and dispersion degree of user information in the Internet of things, possible attacks are detected. Firstly, the average value of the historical user information interval is calculated by power-law weighting method, and the delay interval of the next user information is calculated by exponential distribution function, In reference [6], a new two-way authentication security enhancement protocol was proposed. Different from the traditional RFID authentication protocol, the proposed protocol uses a zero knowledge proof based authentication method to authenticate the membership, The zero knowledge proof is achieved by the real-time information interaction between the prover and the verifier, and the identity security of the participants is defined to the security of their own identity key. In the practical application of traditional access security authentication methods, there is a conflict between the authentication efficiency of handoff between heterogeneous networks and access security, which leads to the security vulnerability of security authentication methods.

Therefore, the multi-source heterogeneous terminal communication module pan-access security authentication technology is proposed to solve the problems mentioned above.

## **2 Design of Universal Access Security Authentication Technology for Multisource Heterogeneous Terminal Communication Module**

### **2.1 Multisource Heterogeneous Data Collection Processing**

In order to process the dynamic flow data collected by multi-source heterogeneous data and be used by other services efficiently, it is necessary to collect multi-source data, which plays an important role in the analysis and decision-making of subsequent services [7]. The collection of heterogeneous data refers to the integration of interconnected distributed heterogeneous data sources to form a "global view" of the underlying data sources, as shown in Fig. 1.



**Fig. 1.** Global view of multi-source heterogeneous data

So that users can easily and transparently access the required data. The process of determining data collection points is as follows:

A circle with radius  $r$  is known, and the following conditions are satisfied:

$$\begin{cases} x_{\max} - r \leq x_s \leq x_{\min} + r \\ y_{\max} - r \leq y_s \leq y_{\min} + r \end{cases} \quad (1)$$

This means that the position of coordinate  $(x_s, y_s)$  will not exceed the square shown in the figure below. If the sides of the square are represented by  $l_1$  and  $l_2$  respectively, the formula is as follows:

$$\begin{cases} l_1 = x_{\min} + r - (x_{\max} - r) = 2r - (x_{\max} - x_{\min}) \\ l_2 = y_{\min} + r - (y_{\max} - r) = 2r - (y_{\max} - y_{\min}) \end{cases} \quad (2)$$

According to the above equation, if the nodes in the circle are more clustered, the length of  $l_1$  and  $l_2$  will increase relatively, and the square area will also increase. On the contrary, if the nodes are more dispersed, the length of  $l_1$  and  $l_2$  and the area of the square will decrease. As can be seen from this, if the range of the square region is large and meets the above conditions, then the mean value of coordinates of all nodes can be used as the coordinates of the collection point. However, if the square range is relatively small, approximate processing should be performed. The square range is divided evenly by using  $l$  root line along the horizontal and vertical directions. By testing each intersecting point, coordinate data satisfying the minimum sum of distance can be obtained, which can be used as the collection point.

In view of the specific situation of heterogeneous data sources and the actual needs of data integration, the data integration solutions are also different. The following is a comparative analysis of the heterogeneous data integration methods commonly used at present [8].

Off-line data integration, off-line data integration The integration of heterogeneous data is achieved by importing and exporting intermediate data files. [9] Its intermediate

files are usually in XML, Excel, Access, JSON and other formats, which are relatively simple to implement. However, due to the use of intermediate files, there are obvious deficiencies in data format conversion and data security, and its synchronization performance is not good [10–13].

**Database Provider Integration.** Database Provider Integration utilizes a mature data provider to access data directly by establishing a connection between heterogeneous data sources. In order to ensure the security of the target system, it is necessary to set up a corresponding view for the required data, and then set up a specific access account to ensure the security of data access.

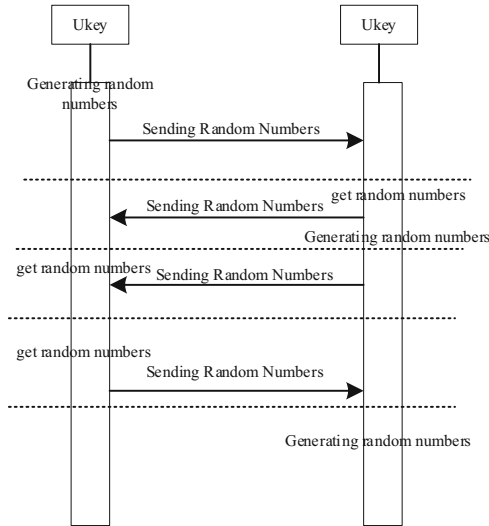
**Data warehouse integration.** Data warehouse integration through the form of data snapshots to heterogeneous data sources copied to the designated data warehouse, so as to achieve effective data integration. This method is usually used by ETL tools to filter the data from the data source periodically, and then load it into the data warehouse for the user to read. During the implementation, the target data can be directly copied to the local database, or copied to the third party temporary table, and then accessed by data interface. The latter is more open and can provide data sources for other systems. Data warehouse integration has both advantages and disadvantages. It has the obvious advantages of fast query and high performance.

**Middleware integration.** Middleware integration is the use of Web services or grid as a building platform tools, and then through XML heterogeneous data conversion and transmission. Web Services are self-describing, self-contained, network-enabled modules that use open XML standards to describe, publish, discover, coordinate, and configure these applications to develop distributed, interoperable applications. Grid is a kind of mechanism that uses the Internet to link all kinds of resources in geography into an organic whole, and can cooperate to accomplish difficult tasks, and provide users with integrated services such as storage and computation. The shared geographical resources include computer system, storage system, communication system, file, database and program, etc. Middleware integration allows consistent access to heterogeneous resources across heterogeneous platforms, providing platform support for large-scale, distributed, heterogeneous data resource access and integration.

In the multi-source and heterogeneous communication module access security authentication, the application scenario is to ensure the access and processing of large-scale heterogeneous data, which is different from the traditional data aggregation, and does not need to access data in the database or other heterogeneous resources, so a unified data communication protocol is adopted to solve the problem of heterogeneous data aggregation.

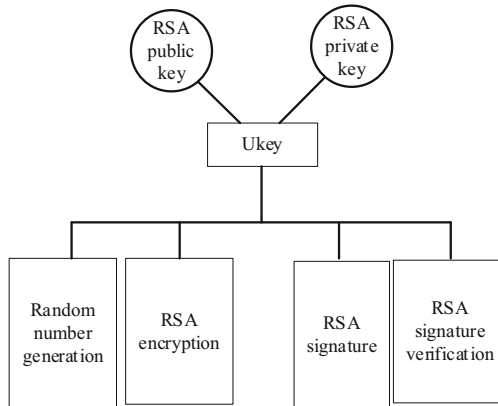
## 2.2 Design Two Way Authentication Architecture

Considering that both sides of the terminal communication module access security authentication in heterogeneous networks are not trusted, there are serious security risks, so a two-way authentication structure is designed. The specific flow of the two-way authentication architecture is shown in Fig. 2.



**Fig. 2.** Flow chart of two way authentication architecture

When the user requests access to the terminal communication module, two-way identity authentication shall be conducted between the two parties. The hardware security authentication module (Ukey) can be inserted on the authentication party and the authentication party. Ukey is a portable plug-in hardware device with the only hardware serial number in the world, as shown in Fig. 3.



**Fig. 3.** Composition of ukey software

Random number generator, RSA encryptor, RSA decryption program, RSA signature program and RSA verification signature program are securely encapsulated, and RSA public and private keys for RSA operation are stored. Ukey also has two types of PIN, the user PIN and the developer PIN. The average user only needs to know the user

PIN, so the user only needs to know the user PIN; as a developer, the developer PIN has more privileges to develop more user-friendly features.

Two-way authentication method is based on RSA encryption and decryption, which needs to use RSA public and private keys. RSA related programs and RSA public and private keys have been securely encapsulated in Ukey, and all of them have been burned and written when Ukey is published to the user. The RSA public and private keys in the same Ukey are not a pair. The certification procedures are as follows:

The user inserts Ukey 1 on the party to be authenticated and confirms that Ukey2 is also inserted on the terminal communication module; the access point establishes a connection application for the access terminal communication module and enters the PIN code to start the two-way authentication. The user PIN with the above PIN of Ukey 1 is designated by the terminal at the time of publishing the Ukey, and the right to modify the user PIN after publishing is reserved to the user; if the user forgets the user PIN, the right to reset the user PIN is reserved to the terminal. The party to be authenticated generates a set of random numbers through the random number generator in Ukey 1. The random numbers are sent to the terminal after being RSA encrypted by the party to be authenticated through Ukey1. After receiving the ciphertext, the terminal decrypts RSA by Ukey2 to get the random number. The terminal signs the random number with RSA through Ukey2 and sends it to the party to be authenticated. After receiving the signature, the terminal verifies the RSA signature through Ukey1. If the verification result is identical to the random number, the authentication is successful and the next step is performed; otherwise, the authentication fails and access is denied.

After successful authentication, the terminal generates a group of random numbers through the random number generating program in Ukey2. The terminal encrypts the random numbers through Ukey2 and sends them to the party to be authenticated after RSA encryption. After the party to be authenticated receives the cipher text, the random numbers are decrypted through Ukey2, and the random numbers are obtained. The party to be authenticated sends the random numbers through Ukey 1 to the terminal after RSA signature, and the party to be authenticated receives the signature and then carries out RSA signature verification through Ukey2. If the verification result is completely consistent with the random number (02), the terminal succeeds in authenticating VD and has access permission. Otherwise, access is denied. The above process needs to be implemented under the support of security authentication protocol.

### 2.3 Determine Access Security Authentication Protocol

When the access party makes an access request, the multi-source heterogeneous terminal communication module initiates  $e_0$  session, generates a random number, and sends authentication requests *query* and  $e_0$ . At the same time, a random number  $e_1$  is generated by the party to be authenticated. After determining the random number, the size of the first eight binary digits of  $e_1$  is calculated, and the authentication threshold  $\chi_1$  is calculated according to  $q(K, e_1)$ . The calculation formula is as follows:



$$\chi_1 = X(q(K, e_1) \oplus e_1 \oplus e_0) \quad (3)$$

The formula above,  $K$  said the identity information in the database, authentication information  $X(K)$  that contains the target object, after the calculation,  $\chi_1$ ,  $e_1$  and  $X(K)$  is sent to the back-end database, the backend database lookup to see if there is meet  $X(K') = X(K)$ , if found no response to record the judge to stop illegal information communication, to find records using the corresponding  $K'$ .

Formula 1 calculates  $\chi'$ . If  $\chi'$  is not equal to  $\chi_1$ , the authentication information is considered to be illegal information and the communication is ended. If it is equal, random number  $e_2$  is generated by the back-end database and  $\chi_2$  is calculated.

$$\chi_2 = X(q(K', e_2) \oplus e_2 \oplus e_0) \quad (4)$$

The result is then sent to the authenticator, and then the data is updated. If  $K' = K$ , then  $K_1$  and  $X(K_1)$  are updated:

$$K_1 = K \oplus e_1 \quad (5)$$

$$X(K_1) = X(K \oplus e_1) \quad (6)$$

If  $K' = K_1$ , we update  $K$ . After receiving the data,  $\chi_2$  is calculated according to Formula 2. By comparing whether  $\chi'_2$  is equal to  $\chi_2$ , if not, the access is judged to be illegal. If equal,  $K$  is updated. The authentication of both parties can be realized through this authentication protocol, and the access security can be further guaranteed.

### 3 Experimental Study on Universal Access Security Authentication of Multisource Heterogeneous Terminal Communication Module

#### 3.1 Experimental Preparation and Design

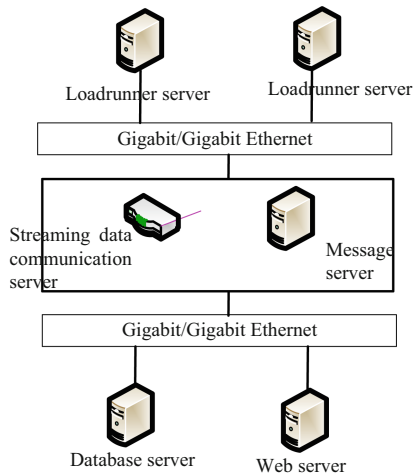
The experiment is based on the platform system which is based on streaming data communication server. The system can monitor the sensor device 24 h, receive a lot of data and view the data in real time. The experimental environment is to deploy web servers, communication servers, message servers, and data storage servers on four servers. The experimental flow is to use Loadrunner to simulate a large number of concurrent scenarios, while simulating heterogeneous multi-source terminal data. After the whole platform system is started up, Loadrunner sends the data to the streaming data communication server, which parses the access data and distributes it to the message server. The message server forwards the message and persists it to the data, where different access security authentication techniques are applied, and the data is finally presented in a web-based application. After the above preparation is completed, the security authentication technology is evaluated from the authentication delay experiment and stability test. Details of the hardware required for the experiment are

shown below, with Table 1 showing the machine configuration. Tools Options Options Options Page.

**Table 1.** Hardware configuration required for experiment

Machine name	Machine configuration	Purpose
m_loadrunner1	Winxp, 4G, Binuclear 15	Data acquisition front end of analog multi-source heterogeneous terminal
m_loadrunner2	Winxp, 4G, Binuclear 15	Analog multi-source heterogeneous terminal information sending front end
m_ccs	Centos6.6, 4G, Binuclear 15	Used as a streaming data communication server
m_activeMQ	Centos6.6, 4G, Binuclear 15	Use as a message server
m_web	Win7, 4G, Binuclear 15	Use as a web server

The experimental environment topology is shown in Fig. 4.

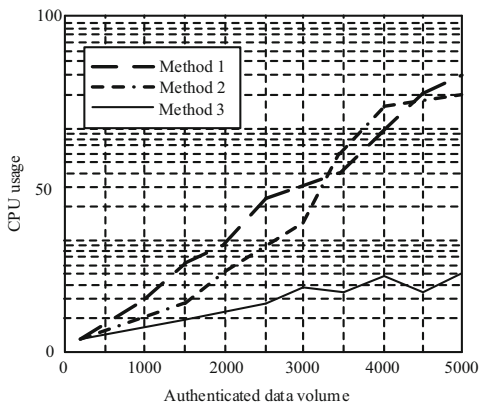


**Fig. 4.** Experimental environment topology

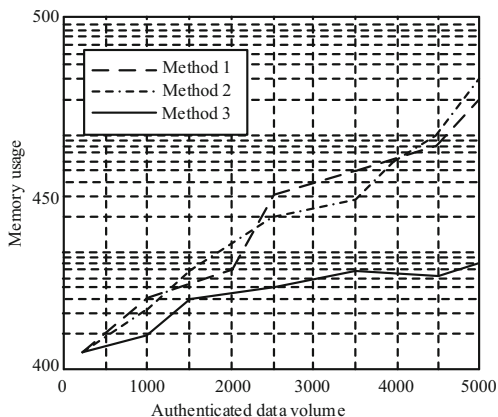
After the preparation of the experiment, the stability and authentication delay of different security authentication technologies are tested.

### 3.2 Security Certification Performance Test

Use LoadRunner to simulate a large number of concurrency and test the stability of security authentication technology. The main observation indexes in the experiment are the data receiving and sending of LoadRunner console, the CPU occupancy rate and memory occupancy rate of streaming data communication server. The experimental results are shown in Fig. 5.



(a) Application of different methods in each security authentication



(b) Memory usage of different methods in each security authentication

**Fig. 5.** Stability test results of the same security authentication technology

In the figure, method 1 represents traditional security authentication technology 1, method 2 represents traditional security authentication method 2, and method 3 represents the proposed security authentication method. In Figure a, it is clear that the CPU

utilization of security authentication technology is increasing with the increase of the amount of authentication data. In contrast, the CPU utilization rate of security authentication technology is always 50%, which shows that the technology runs stably. As can be seen from Figure b, memory usage is also increasing as authentication data increases, and the proposed security authentication technique has lower memory usage compared to the three sets of results, indicating that its support for high concurrency does not depend on memory and does not consume large amounts of memory.

### 3.3 Experimental Results and Analysis of Authentication Delay

In the experiment, different security authentication technologies are used to act on heterogeneous terminals. While data communication is conducted, the authentication delay of different security authentication technologies is monitored. The experimental results are shown in Table 2.

**Table 2.** Test results of authentication delay of different security authentication technologies

Experimental parameters	Traditional security authentication technology 1		Traditional security authentication technology 2		Proposed security authentication technology	
	Non roaming	Roaming	Non roaming	Roaming	Non roaming	Roaming
Certification time	96	638	177	831	86	167
Configure time	15	17	16	17	15	16
Association and handshake time	16	17	15	17	18	19
Switching delay	16	17	15	17	15	14

The unit of data in the table is Ms. the authentication time represents the execution time of the security authentication technology, the configuration time represents the time spent in the pre configuration phase, the association and handshake time represents the time spent in the security association and four handshakes, and the handover delay represents the whole handover time. From the data in the overall observation table, we can see that the time gap between the three security authentication technologies is relatively small in the case of non roaming, and the time difference between the first two security authentication methods is not big in the case of roaming, but it is obvious that the time of the proposed security authentication technology is much less than them. Combined with the stability experiment results of security authentication technology, it can be concluded that the design of multi-source heterogeneous terminal communication module pan access security authentication technology can well overcome the problem of large authentication delay between heterogeneous networks, and ensure the transparency of real-time services and higher security. In practical application, the proposed security authentication technology has more advantages than the other two security authentication technologies.

## 4 Conclusion

With the trend of heterogeneous wireless network convergence becoming more and more obvious, the demand for access security in heterogeneous wireless network is becoming higher and higher. At present, the research on access authentication technology of heterogeneous wireless fusion network is still in the stage of theoretical exploration. The research in this paper can provide unified access authentication technology for heterogeneous wireless fusion network and provide theoretical support for future engineering implementation.

In this paper, multi-source heterogeneous terminal communication module access security authentication is the focus of this research. Based on a large number of research materials and literature, a new authentication method is designed. After the completion of the design, a large number of comparative experiments show that the security authentication technology has better stability and feasibility, and provides better guarantee for multi-source heterogeneous terminal communication security. There are still some difficulties in the design and research. Although some problems in the traditional security authentication technology have been solved, there are still some details, such as access status monitoring, etc. In the future research, this aspect will be studied to further improve the security authentication technology for communication access of multi-source heterogeneous terminals and improve the security of data and access.

## References

1. He, X.: Simulation research on multi source heterogeneous large data cross source scheduling method. *Comput. Simul.* **36**(03), 339–342 (2019)
2. Zheng, X., Ying, Z., Wang, Q., et al.: Research on wireless communication access technology for ubiquitous power Internet of Things. *Electr. Power Constr.* **40**(11), 16–23 (2019)
3. Chen, X., Hu, X., Shen, C., et al.: Research on access authentication technology of power IoT based on Blockchain. *Appl. Electron. Tech.* **45**(11), 77–81 (2019)
4. Yang, C., Jian, Y., Ren, S., et al.: Power LTE network security access technology based on improved authentication protocol. *Electr. Meas. Instrum.* **56**(03), 91–96+102 (2019)
5. Yan, S., Wang, C.: Simulation research on user information authentication security protection of Internet of Things. *Comput. Simul.* **5**, 329–332 (2019)
6. Tan, F.: An improved RFID mutual authentication security hardening protocol. *Control Eng. China* **26**(4), 783–789 (2019)
7. Lu, Y., Zhao, Y., Jiang, L., et al.: Certificateless authentication protocol based on EAP-IBTLS in IoT. *J. Nanjing Univ. Posts Telecommun. (Nat. Sci.)* **39**(01), 62–67 (2019)
8. Tong, S.: Design and verification of security of mobile terminal access workshop information system based on SM4 algorithm. *Mach. Tool Hydraul.* **47**(07), 105–109 (2019)
9. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. *Mobile Netw. Appl.* **26**(5), 1891–1903 (2021). <https://doi.org/10.1007/s11036-020-01725-x>

10. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. *Multimedia Tools Appl.* **76**(4), 5787–5802 (2014). <https://doi.org/10.1007/s11042-014-2408-1>
11. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–17 (2018). <https://doi.org/10.1007/s11036-018-1134-8>
12. Ma, Y., Liu, Z., Wang, Z.: Research on heterogeneous terminal security access technology in edge computing scenario. *Comput. Eng. Appl.* **56**(17), 115–120 (2020)
13. Lin, N., Chen, Z., Zuo, L., et al.: Security analysis and improvement of access protocol for voltage monitoring device in power network. *Comput. Eng. Des.* **40**(11), 3085–3089 (2019)



# Research on Intelligent Operation and Maintenance Technology of Pumped Storage Power Plant Based on 5G

Meng Ye<sup>(✉)</sup>, Guan-jin Huang, Peng Gao, Miao-geng Wang,  
and Xu-hui Zhang

CSG Peak Shaving and Frequency Modulation Power Generation Co., Ltd.,  
Guangzhou 511400, China  
ym42856@163.com

**Abstract.** Due to the poor data communication transmission effect of the current power station operation and maintenance technology, the equipment failure rate is high. To solve this problem, the intelligent operation and maintenance technology of pumped storage power plant based on 5G is designed. The technical design process is completed by constructing the operation state estimation model of pumped storage power plant, setting up the inspection contents and 5G operation and maintenance platform. Through comparison, it can be seen that the application effect of 5G technology is better than current technology.

**Keywords:** 5G network · Pumped storage power plant · Operation and maintenance management · Virtual reality technology

## 1 Introduction

The rapid development of mobile communication industry has led to the rapid development of communication technology, which has brought great convenience to people's communication and provided high quality communication services for people. In recent years, with the improvement of people's requirements on the service quality of the communication industry, 5G mobile communication network technology has achieved rapid development. Compared with 4G communication technology, 5G mobile communication technology brings people a good network experience, greatly improves the network transmission speed, and meets people's needs for high-quality network use. 5G mobile communication itself shows a high data rate, and in the actual use process, it meets the bandwidth requirements of IGHZ [1, 2].

With the continuous progress of society and the enhancement of scientific and technological strength, the development of power grids is also changing with each passing day. Power customers have raised higher standards for the demand for reliability of power supply and the diversification of services. In order to optimize the allocation of resources and adapt to the development process of the current power grids, the State Grid Corporation has put forward a new strategic goal, namely, to build a unified, strong and smart power grid [3, 4]. With a large number of innovative technologies injected into the intelligent construction of power grids, further promote

the development of power plant grid management model by leaps and bounds. Pumped-storage power plant is the core of the power transmission process, which means the operation and maintenance of its intelligent equipment become an important part of the smart grid development strategy.

At present, the installed capacity and power generation capacity of power plants has ranked first in the world. With the progress of power generation technology and the support of power information technology, power plants are moving from centralized control and information management to intelligent, environment-friendly and efficient. Safe, efficient and clean operation is an inevitable trend of the development of power plants. In this context, as a traditional power generation enterprise, in order to ensure production safety, improve management quality and promote energy conservation and consumption reduction, we must make use of digital and information technology to promote the innovative development of enterprises, adjust the production mode and management mode, promote the refined management of power plants, achieve the integration of management and control of the whole life of power plants, ensure the safe production and operation of power generation enterprises and maximize the operating benefits of units in the whole life. With the construction scale and quantity of pumped-storage power station increasing year by year, the problems of construction quality and safety of pumped-storage power station break out frequently, and more and more photovoltaic power stations face the problems of operation and maintenance, design defects, equipment quality defects, construction non-standard and so on, which bring severe challenges to the operation and maintenance of pumped-storage power station. The operation efficiency and management quality of pumped-storage power station will directly affect the operation stability and power generation capacity of the power station. In general, the research and application of operation and maintenance technology of pumped-storage power plants at home and abroad are relatively not diversified in data collection types, especially the serious lack of resource data, detection data and fault data; the functions realized are relatively simple, mainly the viewing, browsing and statistical analysis of the operation information of the equipment of pumped-storage power plants, and the intelligent operation and maintenance function has not been realized yet and needs to be optimized and improved [5, 6].

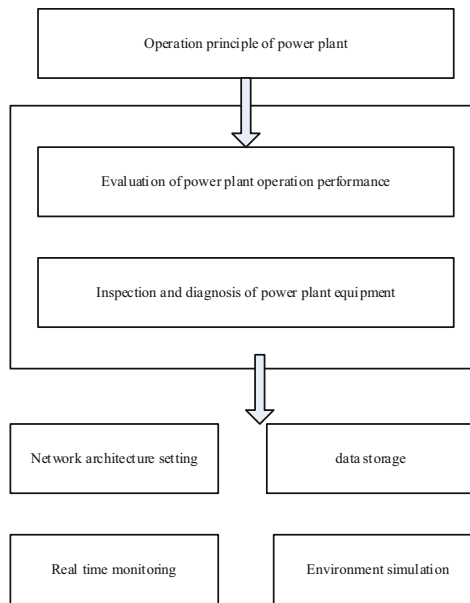
In order to complete the operation data communication of pumped storage power stations with higher efficiency, 5G technology is applied to the operation and maintenance of pumped storage power stations. Therefore, this research uses 5G network as the core technology of intelligent operation and maintenance, and designs the intelligent operation and maintenance technology of pumped storage power plant based on 5G.

## **2 Intelligent Operation and Maintenance Technology Design of Pumped Storage Power Plant Based on 5G**

By studying the related theories of pumped storage power plants, this study analyzes in detail the architecture of “three layers and two networks” in intelligent pumped storage power plants, the characteristics of intelligent equipment, technical characteristics and equipment operation, maintenance and maintenance [7]. Through these analyses, the operation technology, overhaul technology and operation and maintenance management



mode of intelligent pumped storage power plant are studied. In order to meet the higher requirements of pumped storage power plant technology for its management and maintenance, this paper puts forward some suggestions and opinions on the operation and maintenance of pumped storage power plant from the aspects of technical framework, inspection management, maintenance management, equipment risk assessment, defect management, operation and safety control, etc.



**Fig. 1.** Research route of intelligent operation and maintenance technology for pumped storage power plants

In this study, the overall technical design process will be completed according to the content set in the Fig. 1 above. 5G network and virtual reality technology are mainly applied as the core technologies in this research [8]. On the premise of controlling the cost of power plant intelligent operation and maintenance, the application effect of power plant operation and maintenance technology is improved [9]. To provide impetus and technical support for the development of pumped storage power plant.

## 2.1 The Operating State Estimation Model of Pumped Storage Power Plant is Established

In this study, the operating state data of the pumped storage power plant will be collected at first. Data acquisition is the center of data input and output of operation and maintenance management, and also the bridge between operation and maintenance service support platform and other scheduling and control systems [10, 11]. Data collection is also a process of collecting, identifying and selecting data from data

sources. After data collection is completed, it will be used to analyze the operating state of the power plant.

Considering the development trend of the pumped storage power plant degradation in condition assessment is worth more than an absolute standard, according to power plant operation to get the forecast of scores is divided into dynamic and static, in the form of  $a'_{ii}$  and  $a'_{ij}$ , respectively, the two grades respectively according to the proportion of  $\alpha'_{ii}$  and  $\alpha'_{ij}$  coefficients calculated cable one pilot test score information:

$$a_{ii} = a'_{ii}\alpha'_{ii} + a'_{ij}\alpha'_{ij} \quad (1)$$

After the calculation results of single pre-test information are given, all pre-test information  $a_{ij}$  of the power plant shall be calculated according to different importance degrees of each in the state evaluation and different proportions of each information  $\alpha_{ij}$ . Calculate the score of all the pre-test information [12]:

$$b_1 = \frac{\sum_{i=1}^j a_{ii}\alpha_{ii}}{\sum_{i=1}^j a_{ij}} \quad (2)$$

where,  $b_1$  is the score of all the pre-test information.

The scoring principle of on- line monitoring information content can be similar to that of pre-test information content. If there is on- line monitoring information, it can be regarded as some items in the above formula. Since on- line monitoring is not a necessary amount in a state assessment, it defaults if no online information is available at the time of the assessment. Based on the above assumptions, the operational status of the pumped storage power plant can be preliminarily estimated, and the operational status can be analyzed using approximate entropy [13–15] in order to improve the analysis results.

Let the data with length  $n$  be represented by  $B = [b(1), b(2), \dots, b(n)]$ , and the specific algorithm of approximate entropy of data  $b$  is as follows:

The Takens embedding theory is used to construct  $B$  into a group of  $m$ -dimensional vectors, such as:

$$B(i) = [b(i), b(i+1), \dots, b(i+m-1)] \quad (3)$$

The distance between  $B(i)$  and  $B(j)$  is defined as the maximum of the difference between the corresponding elements of two vectors, namely:

$$d[B(i)B(j)] = \max[|b(i+k) - b(j+k)|] \quad (4)$$

Given the threshold value  $u$ , count the number of distances between each row and other rows less than the threshold value  $u$ , i.e.,  $d[B(i)B(j)] < u$ , denoted by  $n_i^m(u)$ , i.e.:

$$n_i^m(u) = \sum_{j=1}^m G(d[B(i)B(j)] - u) \quad (5)$$

where  $G(b)$  is A Heaviside function. By comparing  $n_i^m(u)$  with the total distance  $n - m$ , we get:

$$V_i^m(u) = \frac{n_i^m(u)}{n - m} \tag{6}$$

Take the logarithm sum of  $V_i^m(u)$ , and then take its mean value, which is:

$$\beta^m(u) = \frac{1}{n - m + 1} \sum_{i=1}^{n-m+1} \ln V_i^m(u) \tag{7}$$

Increase the dimension, change the original  $m$  vector into  $m + 1$  dimension, repeat the above several processes to get  $V_i^m(u)$  and  $\beta^{m+1}(u)$ . The approximate entropy of this sequence is obtained:

$$S(m, u, N) = \beta^m(u) - \beta^{m+1}(u) \tag{8}$$

From the above discussion, it can be seen that the approximate entropy value is related to dimension  $m$ , threshold  $u$  and signal length  $n$ . For parameter setting, according to practice Pincus [16, 17],  $m$  is suggested to be 1 and E is suggested to be 0.1–0.2 times of signal standard deviation. In order to ensure the reliability of the data, it is normalized. Normalization is an important method in data preprocessing, that is, the dimensionless data is transformed into dimensionless data. Min-MiAx standardization, also known as deviation standardization, is a linear change of the original data. Where  $B_{\min}$  is the minimum value of sample  $B$  and  $B_{\max}$  is the maximum value of sample. The original value of sample  $B$  is mapped to the value  $B'$  in the interval [0, 1], and its formula is as follows:

$$B' = \frac{B - B_{\min}}{B_{\max} - B_{\min}} \tag{9}$$

According to this formula, the estimated results of operation state are verified, and the verified results are used as the basis of operation and maintenance management.

## 2.2 Power Plant Operation Inspection Content Setting

After analyzing the operation status of the power plant, the patrol inspection technology is used to complete the operation management of the power plant equipment [18]. Through literature study, it can be seen that inspection tours refer to timely recording, judging and reporting the equipment defects or abnormal operating conditions found in the inspection process by regularly or irregularly inspecting the operating conditions and health conditions of various types of equipment in power plants, and handling them in accordance with the relevant provisions, inspecting the existing hidden safety dangers, and notifying the relevant departments to rectify them after discovering them, so as to effectively guarantee the safe and stable operation of the power system through the inspection tours [19].

The periodical inspection mode of pumped-storage power plant can be divided into two categories: normal inspection and overall inspection. Normal patrol refers to the inspection and patrol of the directly observable parts such as the signals of the secondary equipment and the appearance and operation status of the primary equipment, which shall be periodically inspected by the operation and maintenance personnel according to the voltage level of the pumped-storage power plant [20]. In general, the patrol of the pumped-storage power plant shall be conducted once every four days and the patrol of the voltage level equipment with the value of 110 kV or below shall be conducted once every eight days [21]. The comprehensive patrol shall be conducted once every quarter by the relevant personnel of the operation and maintenance department and the transformer substation operation and maintenance room together with the operation and maintenance personnel. Various monitoring equipment are used to monitor the operation condition of the equipment. Carry out seasonal safety inspection; check whether the contents such as false locking, station electricity are in good condition, etc. In this design, the contents of the inspection shall be set as follows (Table 1):

**Table 1.** Inspection content of pumped storage power plant

Involving parts	Project number	The project content
Electrical storage protection device	1	The device and the shell of the collector are clean and complete without damage and well sealed
	2	The indicator lights of the device are normal and correct without flicker
	3	Whether the optical fiber interface of the device is loose
Core control unit	4	The appearance of the device is clean and complete without damage
	5	The secondary terminal connection should be reliable and firm, no loosening off phenomenon
	6	Check whether there is abnormal sound or smell during the operation of the equipment, and whether there is vibration phenomenon
Network signal transceiver device	7	Check the monitoring background computer telemetry data indicating normal, no abnormal alarm signal
	8	Check that all kinds of indicator lights of network switch and its intelligent terminal are normal without flicker and the communication status is normal
	9	Check that the network is working properly

The on-line monitoring device of electrical equipment can collect real-time information of all kinds of equipment for comprehensive processing and analysis, and the calculation results can analyze the trend of equipment status change, evaluate the reliability of equipment, find out the hidden safety trouble existing in equipment as soon as possible, and provide the basis for predicting the maintenance of power grids, which can change the preventive maintenance or fault maintenance into predictive maintenance, reduce the times of power cut of equipment, shorten the time of power cut, and improve the operation and maintenance efficiency of power grids [22].

The on-line monitoring platform of pumped-storage power plant is also an important technical means to realize the state operation and maintenance management of transmission and transformation equipment and enhance the level of lean production and operation management of transmission and transformation specialty. The system realizes real-time sensing, monitoring and early warning, analyzing, diagnosing, evaluating and predicting of all kinds of transmission and transformation equipments by various sensor technology, wide-area communication technology and information processing technology [23]. The construction and popularization of the system have positive and far-reaching significance for improving the intelligence level of power grid and realizing the operation management of transmission and transformation equipments. Therefore, a 5G operation and maintenance platform will be constructed to provide the operation basis for the operation inspection and the operation state estimation model of pumped storage power plant.

### **2.3 5G Operation and Maintenance Platform Construction**

More and more pumped storage power plants entrust professional automation plants to complete the daily operation and maintenance of power plant equipment. These include on-site power plant on duty, provide owners can query the data of the plant anytime, anywhere mobile monitoring client software. With the continuous development of science and technology, looking for more favorable personnel training methods and interactive experience advantages, virtual on-site simulation technology came into being. In order to save cost and concentrate excellent technical backbone to manage more power stations, remote monitoring and unmanned monitoring have developed into an important value-added service item in operation and maintenance system, providing data original collection and storage for each user engineering project and data release service through Internet, and will provide various directional services based on monitoring data in the future, bringing more convenient and more secure monitoring system for users, so as to provide users with satisfactory 7 \* 24-h service. In this study, remote monitoring is equipped with virtual reality simulation technology and mobile phone monitoring, which can quickly complete the operation and maintenance management of power plants. At the same time, we use the network communication platform to complete the round-the-clock remote monitoring of the power plant, based on the digital information platform of water conservancy and hydropower, construct the digital and information management to realize the centralized monitoring and control of several power plants.

In this design, the virtual reality simulation technology will be used as the core technology of the platform. In order to ensure the smooth operation of the platform in

5G network, the virtual reality technology is used to simulate the pumped storage power plant. Virtual reality technology makes use of modern computer technology such as modern computer simulation technology, high sensitivity sensor, artificial intelligence, graphics and image technology and various human-computer interface technology to create human-computer interaction experience with “real scene, real action and real feeling”. In order to apply it to the platform design, we need to keep the safe and stable running state of the management platform, so we need high-quality maintenance and management personnel. Pumped-storage plants involve a wide range of areas, so it is difficult to fully understand a Pumped-storage plant in operation, and it is impossible to realize the familiarity with the functional relationships of various devices and switches by on-site control. Therefore, the use of virtual reality technology in the pumped storage power plant personnel training and field understanding of the situation has a considerable practical significance. The 5G operation and maintenance platform designed in this study can be used to solve the problems of difficult comprehension, monitoring and personnel training, and to optimize the dispatching and personnel training.

The platform uses 3D modeling software to model the whole environment of the hydropower station, and defines the key components such as equipment and switch, etc. By combining the monitoring technology, collecting the field signal, combining the operation situation of the specific power station and the use characteristics, the platform completes the simulation of various operation actions of the valves, buttons and knobs in the operation of the power station, simulates the operation and maintenance of the hydropower station by linkage remote simulation and assists the management, and makes the operation and maintenance of the pumped-storage power station more intuitively. The above design content is connected in order, so far, the intelligent operation and maintenance technology design based on 5G pumped storage power plant is completed.

### **3 Experimental Demonstration and Analysis**

#### **3.1 Experimental Environment Design**

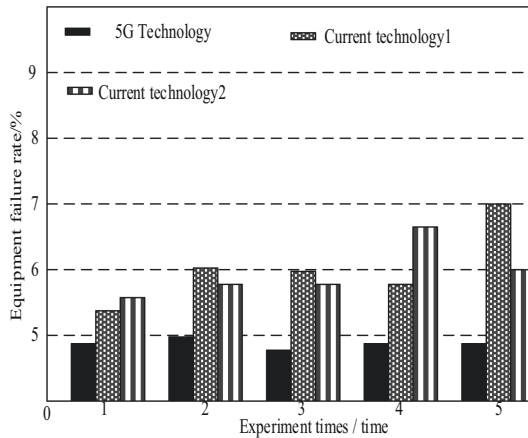
In this study, an intelligent operation and maintenance technology based on 5G is proposed for pumped storage power plants, in order to confirm that this technology has a certain application effect. In this study, the current operation and maintenance technology and 5G operation and maintenance technology are applied to the same pumped storage power plant, corresponding experimental comparison indexes are set, and the application effects of the current technology and 5G operation and maintenance technology are analyzed. In order to avoid the external environment has a certain influence on the experimental results. Set the parameters of the experimental platform in this experiment, and complete the processing process of the experimental platform according to the result of setting the parameters of the platform. 5G technology and current technology will be introduced into this platform to complete the technology application test.

### 3.2 Experimental Control Indexes

In this study, the experimental comparison indexes are set as three parts: the failure rate of power plant equipment, the time of power plant patrol inspection and the time of power plant fault information transmission. During the experiment, the number of experiments was set as 50 times, and a group of average values were taken for every 10 times as the experimental results. According to this average value, the application of 5G technology is analyzed and the comparison process between 5G technology and current technology is completed.

In the collection of experimental data, attention should be paid to the accuracy of experimental data to avoid abnormal experimental results caused by data collection errors. In view of this situation, the experimental data were kept to two decimal places to ensure the authenticity and accuracy of the data (Fig. 2).

### 3.3 Analysis of Experimental Results



**Fig. 2.** Failure rate of power plant equipment

According to the above experimental results, after the use of 5G technology, the failure rate of power plant equipment has changed, showing a trend of gradual decline, indicating that this technology has a better equipment management ability. The current technology is less effective than 5G. After using this technology, the failure rate of the equipment did not change significantly, and the failure rate increased in some tests. Through the analysis, it can be seen that the equipment state estimation module is added in 5G technology, which improves the management ability of power plant equipment to a certain extent. Based on the above analysis results, it can be seen that 5G technology has a high device control capability (Table 2).

**Table 2.** Time of power plant inspection

Method of use	Compare the content of	Specific results/min
5G technology	Maximum patrol time	5.14
	Minimum patrol time	4.74
	Average patrol time	5.07
Current technology 1	Maximum patrol time	6.74
	Minimum patrol time	5.89
	Average patrol time	6.57
Current technology 2	Maximum patrol time	6.15
	Minimum patrol time	5.71
	Average patrol time	6.07

According to the above experimental data, both 5G technology and current technology will consume a lot of time in the inspection of power plants. However, 5G technology has smaller fluctuation and shorter time, which has certain advantages in the management of large power plants. Compared with 5G technology, the current technology takes longer time to conduct inspection on power plants and has greater volatility. The experimental results show that the current technology has poor control over time. Unable to carry out equipment inspection and maintenance in a short time for large power plants (Table 3).

**Table 3.** Time of transmission of fault information in power plant

Method of use	Compare the content of	Specific results/s
5G technology	Maximum message transmission time	10.15
	Minimum information transmission time	8.17
	Average information transfer time	9.54
Current technology 1	Maximum message transmission time	11.25
	Minimum information transmission time	10.58
	Average information transfer time	10.85
Current technology 2	Maximum message transmission time	11.65
	Minimum information transmission time	10.54
	Average information transfer time	11.06

According to the experimental data, after the use of 5G technology, the time of transmission of fault information in power plants has significantly decreased, indicating that 5G technology has improved the intelligence of power plant operation and maintenance management. Compared with 5G technology, the current technology takes a long time in data transmission, which has a certain impact on equipment maintenance. In the future research, this part of performance will be optimized.



The comprehensive analysis of the experimental results of power plant equipment failure rate, power plant inspection time and power plant fault information transmission time shows that the use effect of 5G technology is better. In the future power plant management process, this technology can be used as the main power plant management.

## 4 Conclusion

This paper analyzes the history and present situation of power plant operation and maintenance, and points out the existing problems. Based on the situation that big data and cloud computing technology have been mature and widely used in various industries, this paper studies the intelligent operation and maintenance technology of power plant based on 5G network. In-depth study on the functions of the operation and maintenance platform, further study on VR display technology, enrich the display functions of 5G operation and maintenance technology, and improve the display efficiency and effect.

Compared with 3G and 4G communication technologies, 5G technology has more obvious advantages. As a product of the new era, it is developed on the basis of science and technology, with higher communication security, communication quality and transmission efficiency, and can provide better communication services. In order to promote the application of 5G technology in the intelligent operation and maintenance of pumped storage power stations, it is necessary to increase the full application of high-frequency transmission technology, MIMO technology and multi-carrier technology, clarify the conditions for the intelligent operation and maintenance of pumped storage power stations, and further optimize the operation and maintenance effect. Through this study, it can be seen that the application of 5G technology in the intelligent operation and maintenance of pumped storage power station is feasible and has an ideal application prospect.

## References

1. Zhou, L., Rodrigues, J.J.P.C., Wang, H., Martini, M., Leung, V.C.M.: 5G multimedia communications: theory, technology, and application. *IEEE Multimedia* **26**(1), 8–9 (2019)
2. Kwon, S., Park, S., Cho, H.J., et al.: Towards 5G-based IoT security analysis against Vo5G eavesdropping. *Computing* **45**(3), 1–23 (2021)
3. Tong, D., Hao, F., Wei, C., et al.: Research on construction of data center for intelligent pumped storage power station project. *Hydropower Pumped Storage* **5**(04), 6–10 (2019)
4. Wang, L., Pang, X., Zhu, Y., et al.: Research and application of intelligent maintenance system based on Internet of Things and mobile internet technology. *Electric Power* **52**(03), 177–184 (2019)
5. Yue, Z., Wang, W., Li, D.: Research on augmented reality technology in operation and maintenance of nuclear power plant I&C system. *Ind. Control Comput.* **33**(09), 17–18 (2020)
6. Yuanning, D.E.N.G.: Construction of operation and maintenance control platform of big data transmission network under intelligent background. *Electr. Compon. Inform. Technol.* **3**(09), 106–108 (2019)

7. Zhang, K., Tian, H., Duan, X., et al.: Research on operation and maintenance technology of intelligent substation fault maintenance system. *Adv. Power Syst. Hydroelectr. Eng.* **35**(10), 56–61 (2019)
8. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. *Mob. Netw. Appl.* **26**(5), 1891–1903 (2021). <https://doi.org/10.1007/s11036-020-01725-x>
9. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. *Multimedia Tools Appl.* **76**(4), 5787–5802 (2014)
10. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–17 (2018)
11. Song, Y., Liu, R., Chen, K.: Research on the application of “BIM+” technologies in the operation and management stage of underground pipe gallery. *J. Eng. Manage.* **33**(03), 81–86 (2019)
12. Ye, S., Liu, Q., Chen, M., et al.: Research on management system and construction of operation and maintenance zone. *J. State Grid Technol. Coll.* **22**(01), 40–43 (2019)
13. Chen, M., Zhang, J., Zu, G., et al.: Application of new storage technology in seismic data processing. *Comput. Simul.* **18**(02), 52–58 (2020)
14. Lu, T., Jiang, A., Dong, Y., et al.: Risk assessment of integrated pipeline project operation and maintenance based on information entropy combination weighting-extension model. *J. Water Res. Architect. Eng.* **18**(02), 52–58 (2020)
15. Liu, Y., Hu, Y., Tai, N.: Rule extraction method of operation and maintenance expert system for an intelligent substation based on the decision tree. *J. Electr. Power Sci. Technol.* **34**(01), 123–128 (2019)
16. Tian, F.: Immersive 5G virtual reality visualization display system based on big-data digital city technology. *Math. Probl. Eng.* **2021**(3), 1–9 (2021)
17. Sicari, S., Rizzardi, A., Coen-Porisini, A.: 5G in the Internet of Things era: an overview on security and privacy challenges. *Comput. Netw.* **179**, 107345 (2020)
18. Zhang, S., Wang, Y., Zhou, W.: Towards secure 5G networks: a survey. *Comput. Netw.* **162**, 106871.1-106871.22 (2019)
19. Kaltenberger, F., Silva, A.P., Gosain, A., Wang, L., Nguyen, T.-T.: OpenAirInterface: democratizing Innovation in the 5G Era. *Computer Netw.* **176**, 284–293 (2020)
20. Subramanya, T., Harutyunyan, D., Riggio, R.: Machine learning-driven service function chain placement and scaling in MEC-enabled 5G networks. *Comput. Netw.* **166**, 106980-1-106980-16 (2020)
21. Rocchetta, R., Bellani, L., Compare, M., et al.: A reinforcement learning framework for optimal operation and maintenance of power grids. *Appl. Energy* **241**(15), 291–301 (2019)
22. Yildirim, M., Gebraeel, N.Z., Sun, X.A.: Leveraging predictive analytics to control and coordinate operations, asset loading, and maintenance. *IEEE Trans. Power Syst.* **34**(6), 4279–4290 (2019)
23. Nico, V., Rodriguez, J.C., Punch, J.: A vibration energy harvester and power management solution for battery-free operation of wireless sensor nodes. *Sensors* **19**(17), 3776–3785 (2019)



# Smart Grid Service Transmission Accuracy Optimization Technology Based on 5G Technology

Li-kuan Gong<sup>1(✉)</sup>, Jian-yong Zhou<sup>1</sup>, Guo-yi Zhang<sup>2</sup>, Tong-hao Wu<sup>1</sup>,  
and Zhi-wei Liu<sup>1</sup>

<sup>1</sup> Shenzhen Power Supply Bureau Power Dispatching and Control Center,  
Shenzhen 440304, China

<sup>2</sup> CSG Power Dispatching and Control Center, Guangzhou 510670, China

**Abstract.** Smart grid needs to have high stable transmission control accuracy, fast dynamic response performance and strong anti-interference ability. Conventional technology transmission of smart grid services, data transmission delay is long, in order to solve this problem, the design of smart grid service transmission accuracy optimization technology based on 5G technology. Build the general structure of the smart grid business transmission, define the business end of communication, specification data storage format, optimized algorithm for data transmission retreat, build 5 g network slice application scenarios, meet the requirement of smart grid, all kinds of business, and according to the business terminal sends commands, identify business communication resources, control of business information data transmission. The experimental results show that compared with the conventional technology, the design technology can reduce the waiting time and execution time of power network service transmission, and shorten the delay of data transmission.

**Keywords:** 5G technology · Smart grid · Service transmission · Communication technology

## 1 Introduction

The State Grid Corporation of China has taken the construction of world-class energy Internet enterprises as the construction goal, and continuously improved the construction level of information and communication networks. With the deepening of the construction, various business departments have built up some typical applications of the Internet of Things based on the actual needs through the use of terminals and networks. The number of terminals serving the production business of power grids has reached more than one billion, and the power Internet of Things has taken shape [1]. However, during the construction of the Internet of Things, the gradually exposed deficiencies have not been completely solved.

In reference [2], in order to solve the control parameter optimization problem of hybrid HVDC system, a new meta heuristic algorithm, whale algorithm, is introduced. Aiming at the problem that the global search ability of the algorithm is poor, the hybrid optimization theory is introduced and improved by using simulated annealing and

League selection mechanism. The improved algorithm effectively balances the global search and local development ability. Taking the bipolar two terminal hybrid HVDC test system as the research object, the improved whale algorithm is used to optimize the proportional integral (PI) parameters in the rectifier side constant current control and the PI parameters of the inverter side DQ axis double loop controller, and the algorithm is realized by the joint simulation of MATLAB and PSCAD software. In reference [3], a double closed-loop control method based on network current feedback is proposed. The control method based on pole assignment is used to improve the inverter instantaneous voltage PID controller. The simulation results show that the PID control method based on pole assignment makes the inverter dynamic response fast and output voltage thd low. The large design method based on pole assignment is applied to the inverter voltage current double loop control system. In reference [4], an optimal configuration method of flexible AC transmission equipment based on adaptive particle swarm optimization is proposed. The optimal configuration method takes the optimal voltage stability, maximum available transmission capacity and minimum comprehensive cost as the objectives, establishes the optimization model, and uses the adaptive particle swarm optimization algorithm based on the improvement of distribution entropy to determine the optimal configuration scheme. However, the above traditional methods all have the problems of long waiting time and execution time of power grid service transmission, resulting in long delay of data transmission.

In order to solve the above problems and provide personalized and deterministic service support for different services, we shall, on the basis of the study of typical smart grid business scenarios of 5G, focus on tackling the collaborative and efficient transmission technology of ubiquitous service end networks of smart grids based on 5G so as to form effective solutions and relevant research results. Meanwhile, we shall pull through operators to verify the relevant business, explore the feasibility of 5G in the power industry and lay a foundation for the large-scale application of 5G. 5G has three application scenarios: ultra-high bandwidth, high reliability and ultra-low delay, and ultra-large scale connection. It can comprehensively improve the operation efficiency and intelligent level of traditional vertical industries, and is highly compatible with the needs of smart grid communication business. The agility and customizability of 5G network slicing technology can perfectly connect and fully meet the requirements of diversity and isolation of grid business, and create exclusive networks for different business needs of smart grid.

## **2 Optimization Technology Design of Smart Grid Service Transmission Accuracy Based on 5G Technology**

### **2.1 Build the Overall Structure of Smart Grid Service Transmission**

The overall structure of smart grid service transmission is constructed, and the service terminal communication method is defined. The overall structure of power grid service transmission is composed of five functional modules, namely, data collection module, information data control module, public cloud platform, terminal and network communication. The data collection module is used to collect the information data for

transmission in real time, send the information data to the control unit, and use the control module to store the received information data. The control unit is made up of MicroinoduCore with the size of  $26.36 \text{ mm} \times 28.93 \text{ mm} \times 31.53 \text{ mm}$ , with the ATMEGA328P microcontroller as the core chip, powered by MICRO and by USB serial communication, so that the control unit has a compatible and standard peripheral interface and responds to the interactive commands sent by the cloud platform. The network communication module is composed of ENC28J70 Ethernet transceiver chip and RJ63 interface, which provides support for real-time communication between cloud platform and terminal, converts transmitted information data into corresponding information interaction commands, and business terminals download and receive information data to realize information interaction.

The IEC 61850 communication mode of the service terminal is defined and applied to the overall structure of the service transmission of the power grid. The communication structure of IEC 61850 is divided into four layers: physical layer, logical layer, data attribute layer and data object layer. Each layer is equipped with several logical nodes, and all nodes are closely associated with each other, which are composed of data objects for information exchange and cooperate with the task of business transmission. When constructing the IEC61850 distributed communication structure, the parameters and attributes of the information data are extracted firstly, and the information data of the logical nodes are combined according to certain granularity rules. Then self-description configuration IEC 61850 file, hosting the overall structure of the business transmission specific extensible markup language, to achieve the transfer of data information, the configuration file will be divided into IED capability description file, specification description file, terminal configuration description file, CID description file 4 formats, in which IED and CID file describe the communication network topology connection, indicating the information template has not been instantiated. According to the logic node of DG, it provides support for data transmission, realizes the automatic description of DG, and further exports the instantiation communication file of business terminal. Finally, mapping IEC 61850 communication service to communication protocol stack, defining communication service interface, adopting web service, GOOS service, MMS protocol, IEC 60870-5-104 protocol, aggregating data through protocol stack, specifying information data transmission mode, including MMS protocol, sampling value, Ethernet type of message mapping, etc., to realize the independence of communication protocol of business information data. Thus the overall structure of smart grid service transmission is constructed.

## 2.2 Optimize the Backoff Algorithm of Power Grid Business Data Transmission

When the information and data communication is realized at the power grid business terminal, the backoff algorithm of the interaction mechanism is optimized to ensure the integrity of information and data transmission. The embedded development board is used as the main control component of the power grid business aggregation terminal, and the S5PV210 chip is used as the core chip of the development board to store the communication resources of the power grid business [5]. The storage format set is shown in the following table (Table 1):

**Table 1.** Storage format of power grid service communication resources

The name of the data	The data type	The length of the data	Instructions
Id	Varchar	8	Site longitude
Address	Int	13	Site latitude
Message	Varchar	50	The site name
Name	Varchar	20	Site number
Latitude	Int	4	Contact
Longitude	Int	50	Site type
Phone	Varchar	4	Equipment serial number
EquipmentName	Varchar	4	Operating state of equipment
Type	Int	30	Subordinate to the site
Thor	Int	10	Administrator number
EquipmentNO	Varchar	50	Administrator privileges
EquipmentStatus	Varchar	8	Note

The service aggregation terminal is used as client and the management and control center is used as server to make the aggregation terminal communicate with the management and control center through socket. The communication mode follows the C/S structure and monitors the service request of the control center in real time. When the Management and Control Center sends the grid business, it shall create the socket original socket, intercept the data business of the Management and Control Center and copy it to the buffer. When the data request reaches a certain number, it shall verify the data. When the data is correct, it shall be transmitted to the data processing thread. Otherwise, the data shall be discarded to the socket data monitoring. When a data processing thread processes the requested data, it shall remove the data part of the Ethernet frame, determine the data type actually received, determine the data source, and then bind a specific socket according to the list of power grid business resources, adopt a custom data format, check the MAC address, send the business request to the network port, and the business aggregation terminal shall parse the data request of the communication resource management thread, and respond to the business instructions of the power grid control center. After access to IEC 61850 communication technology, the service terminal is regarded as the destination node, and the terminal receives the data packets sent by the source node, replies the ACK data frame to the source node. If the information exchange fails, the access to the wireless channel media shall be stopped, the source node sending the data shall enter the competition period again, the backoff algorithm shall be implemented, the number of retransmissions of the information packets shall be increased, and the time window for competition of all source nodes shall be increased exponentially. In the process of competing for data transmission channels, the source node shall reduce the probability of successful competition of the node sending the data, increase the probability of successful competition of the node sending the data, and ensure that the information data can be successfully connected to the transmission channel [6]. In order to reduce the probability of conflict in the subsequent communication and reduce the possibility of conflict

again, the backoff time of the counter is calculated by the product of the length of the time slot and a random number.

The calculation formula of the execution function  $F$  of the retreat counter is as follows:

$$F = \xi \text{Min}(2Q, Q_{\max}) \quad (1)$$

Where  $Q$  is the competitive window value of nodes,  $Q_{\max}$  is the maximum competitive window value, and  $\xi$  is the random number of  $[0, Q]$ . The idle time of the channel is detected. When the time is greater than the set threshold, the backoff counter is decremented by 1 by executing the function until the backoff count is decremented to 0, and the information data is sent again until all the information data are sent to the channel. So far, the backoff algorithm of power grid business data transmission is optimized.

### 2.3 Build 5G Network Slice Application Network Scenario

Application of 5g network slicing application network scenarios, so that business data transmission can meet the needs of power grid communication function.

#### Analysis of 5G End to End Network Slicing System

Combined with 5G network slicing mode, the 5G end-to-end network slicing system is analyzed. 5 G network chip overall architecture includes an integrated integration system of cloud, pipeline, terminal and security. In particular, cloud provides users in the power industry with the ability to build a more open, convenient and self-management platform, and cloud platforms for smart grid applications can be divided into two categories: the first category is the traditional power business platform, corresponding to the power generation control area; and the second category is the communication management support platform, including terminal management, business management, chip management, information management, dispatch management and other modules, corresponding to the power management information area. Pipelines, namely 5G communication networks, include wireless base stations, transmission networks, core networks, etc., jointly provide slicing services for smart grids, and provide exclusive sub-slicing services for different subdivisions of traditional power services, so as to ensure the safety isolation of power services and the requirements of network indicators. Reliable connection between power terminals and master stations shall be realized through the connection between various power service platforms and modules, and operators' networks shall open interfaces through network capacity to realize the opening and sharing of terminals and network information, thereby realizing the secondary operation of network chips [7–9]. Terminals refer to remote equipment connected to communication networks, mainly including intelligent distributed distribution automation terminals, intelligent ammeters, high-definition cameras, unmanned aerial vehicles, patrol robots and other special power terminals. The requirements of the terminals for network bandwidth, delay, reliability, number of connections and other requirements correspond to the three major application scenarios of 5G network slicing, and the security isolation between services is realized through

slicing technologies. The security system covers three levels, namely, cloud, pipeline and terminal. Before the power-generating business and control business are accessed to the cloud platform, physical and logical isolation shall be conducted according to the security requirements of the State and the power industry. The security mechanism of 5G smart grid shall focus on pipelines and terminals, and provide unified authentication and authentication framework, multi-level network slice security management, secondary authentication and encryption, network domain security, border firewall, etc. through the 5G network to enhance the access security of pipeline networks and terminals. Thus the analysis of 5G end-to-end network slicing system is completed.

#### **2.4 Application of 5G Network Slice in Intelligent Network**

The 5G network slicing technology is used as an entry point for operators to assist the construction of smart grid, and the network is customized according to the characteristics of grid services to meet the requirements of various smart grid services [10, 11]. Application of distributed power distribution automation, distributed power distribution automation is a comprehensive information management platform integrating computer technology, data transmission, control technology, modern equipment and management, which protects and controls the distribution network, detects the status information of distribution network lines and equipment through automatic relay protection devices, realizes intelligent judgment, analysis, fault location, fault isolation, power supply recovery in non-fault areas and other operations, realizes fault diagnosis and accurate positioning of distribution network lines and network equipment, rapidly isolates distribution network fault sections and fault equipment, and minimizes the time and scope of power outage due to faults, so as to reduce distribution network fault handling time from the minute level to the millisecond level. Precise load control, which is an important technical guarantee for precise excision of interruptible load. In the case of grid failure or unbalance of power consumption, load control will reduce load, improve the economy and security of grid operation, and enhance the investment benefits of grid enterprises by means of power stability control, emergency excision of load and low-frequency and low-voltage load shedding devices [12]. Establish a communication management support platform, provide an open interface for operating capacity, provide the traditional electric power business platform with terminal status, monitoring of business status and chip management, and achieve the manageability and controllability of electric power communication [13, 14]. Meanwhile, the communication management support platform can collect and summarize the terminal, business and network data of the traditional electric power business platform, and provide big data analysis and other advanced application functions. At this point, the intelligent network, 5G network slicing application network scene construction and application.

#### **2.5 Transmission and Control of Smart Grid Business Data**

According to the actual content of the instructions sent by the service terminal, it identifies the service communication resources of the power network and realizes the management and control of information and data transmission [15]. The server that processes business communication resources in different regions of power communication shall be



selected, and be equivalent to a virtual machine, so that the servers in the same region are located in the same local area network, the server of power communication is represented by numbers, the processing speed of the server for business resources is unified, and different virtual machine control tasks are assigned according to the identified communication resources [16]. Dynamic migration is adopted in the resource transfer mode. Firstly, the monitoring node load of the terminal is determined, and the physical nodes that need to be migrated are identified. Set a fixed threshold, and take CPU utilization as a criterion. When the node utilization exceeds the threshold, move out the appropriate virtual machine, that is, move out the communication resources processed by the server, put the resources on other nodes, and reduce the load of the node. When the node utilization is lower than the set threshold, move out all the virtual machines run by the node, and eliminate the business resources run by the node, so as to ensure the quality of power communication services [17, 18]. The smoothing index of time series is used to predict the CPU utilization of different time windows of nodes. The calculation formula is as follows:

$$Q = \alpha x_t + \alpha x_{t-1} + \cdots + \alpha^{n+1} x_{t-1} + \alpha_t \quad (2)$$

In the formula,  $Q$  is the predicted value of CPU utilization,  $\alpha$  is the prediction parameter less than 1,  $\phi$  is a random variable with smooth exponential normal distribution,  $n$  is the time change of control time window, and  $x_t$  is the historical load weight of node at  $t$ . Through formula (2), the power communication node to be migrated is determined, the business resource data stored in the server is copied, the volume of virtual machine is measured, and the virtual machine with the smallest volume is selected for dynamic migration. The calculation formula of volume  $V$  is as follows:

$$V = \frac{1}{1 - V_1} \times \frac{1}{1 - V_2} \times \frac{1}{1 - V_3} \quad (3)$$

In the formula,  $V_1$  is the CPU utilization rate of the virtual machine,  $V_2$  is the memory utilization rate of the virtual machine, and  $V_3$  is the hard disk utilization rate of the virtual machine [19]. Through the above three indicators, the utilization rate of business resources is determined. When the utilization rate is lower, the volume of the virtual machine is smaller. Determine the physical nodes and virtual machines that need to be scheduled, use the cloud computing scheduling algorithm to realize the dynamic migration of commercial resources, count the subtask completion time, computing performance and real-time network bandwidth of all virtual machines, integrate the power communication lines into the cloud, take the scheduling node as the master node, the scheduling task as the cloud task, and other physical nodes as the slave nodes [20]. So that each cloud task can only be executed on one commercial resource, the cloud tasks are numbered, and the business transmission tasks are sorted from high to low according to the priority. Set the constraint conditions of the transmission task, set the master node as  $i$  and the slave node as  $j$ , then the time constraint function formula is:

$$P = A(i) + B(j) + C(i, j) \quad (4)$$

In the formula,  $P$  is the communication time from the master node to the slave node for cloud tasks,  $A(i)$  is the time required for the commercial resource transmission of the master node  $i$ ,  $B(j)$  is the time required for the resource reception of the slave node  $e$ , and  $C(i, j)$  is the time required for the resource data to be transmitted from the node  $i$  router to the node  $j$  router through Ethernet [10, 11]. The formula of reliability constraint function is as follows [21]:

$$D = \frac{\sum_{k=1}^O M_k}{\xi} \quad (5)$$

In the formula,  $D$  is the reliability of resource scheduling scheme,  $\xi$  is the number of business resources running by virtual machine,  $O$  is the total number of cloud tasks, and  $M_k$  is the possibility of commercial resources selection for the  $k$  cloud task after resource scheduling. The communication time  $P$  and reliability  $D$  of the main node  $i$  to slave node  $j$ . The optimal scheduling link is selected to allocate the traffic transmission task to the link road, and the inherent time of the virtual machine is updated [22, 23]. After the scheduling is completed, the effective processing of node  $j$  to the transmission task is calculated, the formula is:

$$\zeta = \frac{L(j)}{\beta} \quad (6)$$

In the formula,  $\zeta$  is the processing capacity of node  $j$  for commercial resources, that is, the feasibility of downloading business resources,  $L(j)$  is the computing length of cloud task of virtual machine, and  $\beta$  is the download time period of cloud task, which is the difference between the download completion time and the download start time. Adjust the calculation length of the transmission task in time to make the prediction value  $\zeta$  of the effective processing reach the maximum, so as to replace the original performance parameters of the virtual machine. When the transmission task is completed, the next transmission task is assigned in real time. According to the parameter changes of the virtual machine, the priority of the transmission task is adjusted at any time to provide more reliable business resources until all virtual machines are traversed and all power is allocated Network service transmission task. So far, the transmission control of smart grid business data has been completed, and the technical design of smart grid business transmission accuracy optimization based on 5g technology has been realized.

### 3 Analysis of Experimental Demonstration

The optimization technology is recorded as experimental group A, and the two commonly used optimization technologies of smart grid service transmission accuracy are recorded as experimental group B and experimental group C. the smart grid service is transmitted respectively, and the data transmission delay of the three technologies is compared.

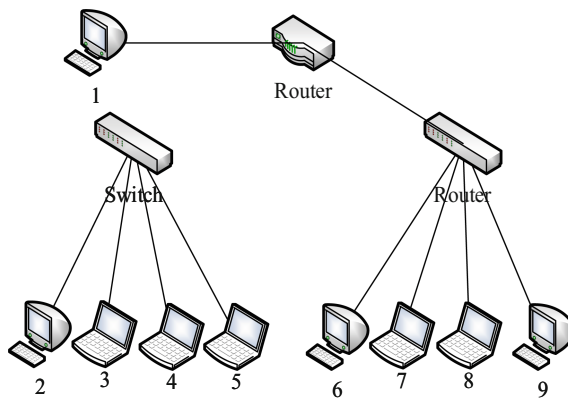
### 3.1 Experimental Preparation

The power simulation platform is built by using cloudsim simulation software. The infrastructure layer consists of 1000 physical nodes, each node has 55 generators and 189 lines. The power grid scale is medium, and the host configuration is shown in Table 2.

**Table 2.** Host configuration of power grid simulation environment

The physical machine	Memory	CPU	Resource read speed (MB/s)	The response time (ms)
1	16 GB	Single-core 2.4 GHZ	44.3	13.3
2	16 GB	Binuclear 2 GHZ	50.7	15.3
3	8 GB	Single-core 1.8 GHZ	55.1	17.9
4	16 GB	Single-core 2.67 GHZ	53.9	16.4
5	8 GB	Four core 2.1 GHZ	49.8	15.2
6	16 GB	Binuclear 2 GHZ	55.4	17.8
7	8 GB	Single-core 2.4 GHZ	49.2	14.2
8	8 GB	Single-core 2.8 GHZ	49.8	13.1
9	16 GB	Binuclear 1.8 GHZ	53.2	15.8

Multiple frames are formed into a cluster to interconnect the host server through the switch, so that the test results of the three technologies are closer to the practical application. The communication network topology is shown in Fig. 1:



**Fig. 1.** Topological structure of power grid communication network

As shown in Fig. 1, host 1 serves as the control node, host 2 serves as the scheduling controller and cluster controller, and the other hosts serve as node controllers to provide communication resources for power grid business. The wireless local

area network (WLAN) with IP address of 187.132.6.2, network type of 5G and public network interface of WLAN0 was selected for the experiment. The simulation platform has A total of 5 stations. Station A is the communication field station of power communication equipment A, Station B and Station C are optical cable jumping stations, Station D is the communication master station of communication equipment B, and Station S is the core master station.

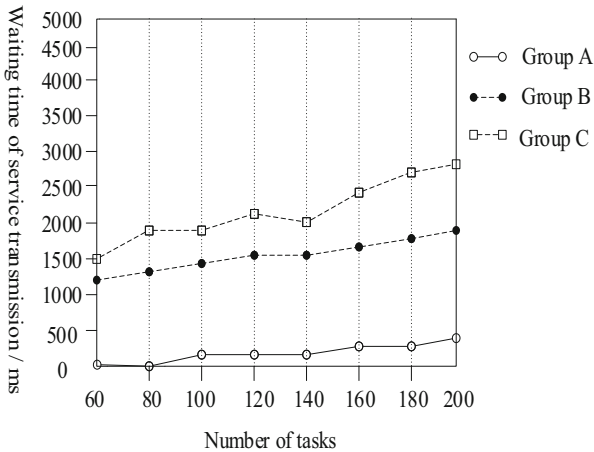
Simulation parameters of 5G communication network are shown in Table 3:

**Table 3.** Setting of communication network simulation parameters

Parameter	Numerical	Parameter	Numerical
Simulation scene size	900 m × 900 m	Channel type	WirelessChannel
The simulation time	20 min	Channel bandwidth	15 MHz
Working frequency	6.837 GHz	Routing type	AODV
Network Interface Type	WirelessPhyExt	Interface queue type	PriQueue
Wireless transmission type	TwoRayGmund	Type of antenna	OmmiAntenna
MAC protocols	Mac802-11p	SIFS	30 μs
DIFS	40 μs	SLOT	15 μs
Minimum competition window	20	Node running speed	35/s

### 3.2 The Experimental Results

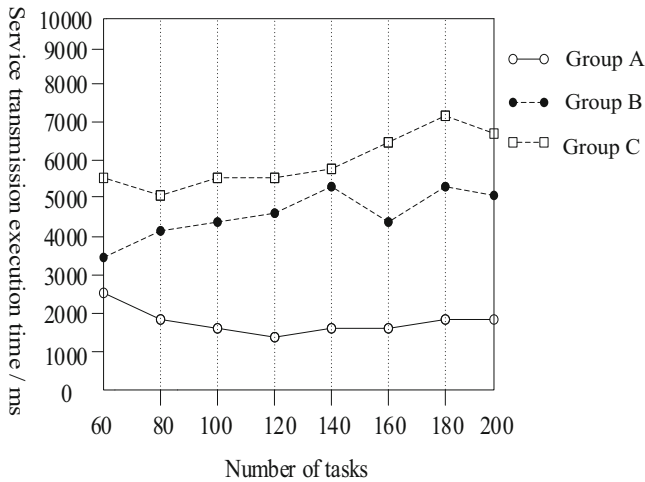
In stand-alone mode, through the host 1 business running power grid transmission task, business transmission resource set, composed of 60–200 independent tasks, task span of 20, change the number of task resource scheduling, optimization technology application three sets of grid services transmission precision, more business transmission grid average waiting time. In the process of smart grid business execution, the current common service waiting time is 1500 ms–2500 ms. The experiment result is shown in Fig. 2:



**Fig. 2.** Comparison results of waiting time for service transmission

The Fig. 2 shows that the experimental group A grid business transmission of average waiting time of 113 m, and the number of tasks less than 80, resource scheduling without waiting time, the experimental group B average waiting time of 1504 m, the experimental group C average waiting time for 2045 ms, compared with the experiment group B and group C, group A waiting time decreased by 1391 ms and 1392 ms respectively.

On the basis of a group of experiments, the task with high trust was selected to match with the communication resources of the power grid business, and the execution time of the three groups of technology transmission business was counted. According to the current survey of smart grid services, the average execution time of single service transmission is 3000 ms–6000 ms.



**Fig. 3.** Comparison results of service transmission execution time

As can be seen from Fig. 3, the average execution time of power network service transmission in experimental group A is 1826 ms, the average execution time of experimental group B is 4467 ms, and the average execution time of experimental group C is 6078 ms. Compared with experimental group B and experimental group C, the execution time of experimental group A is reduced by 2641 ms and 4252 ms respectively. To sum up, compared with the two groups of commonly used technologies, this design technology reduces the waiting time and execution time of power grid business transmission, shortens the data transmission delay of power grid business, and the transmission precision optimization effect of smart grid business is better.

## 4 Conclusion

The design technology gives full play to the communication advantages of 5G technology and shortens the data transmission delay of power grid business. However, there are still some shortcomings in this study. In future studies, in-chip BRAM will be adopted to make full use of digital front-end resources, improve the link transmission rate, expand the data cache space, and support larger scale power grid service transmission.

## References

1. Liu, G., Wang, X., Huang, J., et al.: Analysis of performance for information of distribution network automation under different transmission protocols. *Electr. Meas. Instrum.* **57**(17), 99–105+146 (2020)
2. Liu, Q., Tang, X., Yang, J., et al.: PI parameters optimization based on improved whales optimization algorithm for hybrid high voltage direct current system. *Electr. Power Constr.* **40**(04), 30–37 (2019)
3. Yu, Z., Wang, W.: Optimal control of inverter power supply quality. *Comput. Simul.* **36**(04), 63–66+134 (2019)
4. Qi, W., Wang, Q., Chen, Q., et al.: FACTS optimal configuration of high proportion photovoltaic power grid considering voltage stability. *Renew. Energy Resour.* **12**, 1786–1793 (2019)
5. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. *Mob. Netw. Appl.* **26**(5), 1891–1903 (2021). <https://doi.org/10.1007/s11036-020-01725-x>
6. Lu, Z., Zhang, Y., Zhao, Y., et al.: A method to improve the reliability of relay protection service transmission channels. *J. State Grid Technol. Coll.* **23**(5), 33–37 (2020)
7. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. *Multimedia Tools Appl.* **76**(4), 5787–5802 (2014). <https://doi.org/10.1007/s11042-014-2408-1>
8. Lv, Y., Yang, Y., Dong, Y., et al.: Application of 5G to current differential protection of distribution network. *Telecommun. Sci.* **36**(2), 83–89 (2020)
9. Li, W., Fan, H., Shao, H., et al.: Research on transmission evolution strategy based on packet enhanced OTN in power communication network. *Electr. Power Inf. Commun. Technol.* **18**(9), 64–69 (2020)
10. Li, Y., Zhu, W.: Modern optical transmission network technology and its application in power transmission network. *Commun. Technol.* **53**(6), 1569–1574 (2020)
11. Yan, L., Chen, Z., Yu, X., et al.: Security interaction framework for electricity service in new-type town based on quantum key distribution. *Autom. Electr. Power Syst.* **44**(8), 28–35 (2020)
12. Tang, X., Liu, Q., Sun, C., et al.: Study on network structure and information transmission optimization of secondary system in smart substation. *Power Syst. Big Data* **23**(3), 77–84 (2020)
13. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. *Mob. Net. Appl.* **24**(1), 5–17 (2018). <https://doi.org/10.1007/s11036-018-1134-8>
14. Yang, G.: TD-LTE specialized network for distribution network service transmission performance test and analysis. *Comput. Technol. Autom.* **38**(4), 66–69 (2019)

15. Gao, Z., Zhao, S., Ma, Y., et al.: International business data transmission scheme based on MDI-QKD protocol. *Chin. J. Quant. Electron.* **36**(1), 34–39 (2019)
16. Zhang, Z., Li, L.: High-voltage DC high-efficiency detection and simulation of power transmission and distribution in communication system. *Comput. Simul.* **37**(4), 169–172 (2020)
17. Zhou, Y.: Research on the improvement strategy of grid user service satisfaction. *Adv. Appl. Math.* **08**(2), 242–249 (2019)
18. Shang, K.: Semantic - based service discovery in grid environment. *J. Intell. Fuzzy Syst.* **39**(1), 1–10 (2020)
19. Zhang, K., Troitzsch, S., Hanif, S., et al.: Coordinated market design for peer-to-peer energy trade and ancillary services in distribution grids. *IEEE Trans. Smart Grid* **11**(4), 2929–2941 (2020)
20. Liu, J.N., Weng, J., Yang, A., et al.: Enabling efficient and privacy-preserving aggregation communication and function query for fog computing based smart grid. *IEEE Trans. Smart Grid* **11**(1), 247–257 (2019)
21. Milievi, L.: An improved upper bound for the grid Ramsey problem. *J. Graph Theory* **94**(4), 509–517 (2020)
22. Hasan, M.M., et al.: Cloud-centric collaborative security service placement for advanced metering infrastructures. *IEEE Trans. Smart Grid* **10**(2), 1339–1348 (2019)
23. Jiang, W., Fang, X., Ding, J.: Gaussian kernel fuzzy C-means algorithm for service resource allocation. *Sci. Program.* **2020**(3), 1–6 (2020)



# Research on Hybrid Communication Networking Protocol Optimization Technology of Ubiquitous Electric Internet of Things

Bao-ren Chen<sup>1(✉)</sup>, Dan-ke Hong<sup>1</sup>, Li Wang<sup>1</sup>, Yong-tong Ou<sup>2</sup>,  
and Xin-hui Zhong<sup>2</sup>

<sup>1</sup> Communication Office of China Southern Power Grid,  
Guangzhou 510670, China  
cbr42561@163.com

<sup>2</sup> Digital Power Grid Branch of China Southern Power Grid Digital Power Grid  
Research Institute, Guangzhou 510670, China

**Abstract.** In order to solve the problem of high routing overhead caused by the lack of relay coding in hybrid communication network protocols, this paper studies the optimization technology of hybrid communication network protocols. According to the bandwidth requirement, delay requirement and reliability requirement of communication network, the requirement characteristics of communication network are extracted. Through the establishment of multi-frequency node network, configuration of node white list and agent node selection, the design of hybrid communication network architecture is completed. Design communication packet routing algorithm, according to a certain topology and logical relationship to organize the network nodes. Based on the establishment of hybrid communication network architecture and the design of routing algorithm, the ubiquitous hybrid communication network protocol is optimized. Experimental results show that the total routing overhead of the proposed protocol is 284 packets, 920 packets and 548 packets less than that of the existing protocols, respectively.

**Keywords:** Ubiquitous electric Internet of Things · Hybrid communication · Hybrid networking · Communication protocol

## 1 Introduction

At present, major countries in the world have taken the Internet of Things as a major strategy to seize the commanding heights of a new round of economic and scientific and technological development, and China has also taken the Internet of Things as a strategic emerging industry that has risen to the height of national development priorities, and explicitly proposed in the Outline of the 12th Five-Year Plan to promote the research on key technologies of the Internet of Things and the application demonstration in key fields, which has become an important content of the “IOT +” national action plan for development in recent years [1].

Ubiquitous power IOT communication services are characterized by diversification and wide coverage, covering power transmission, transformation, power generation,



dispatch and other power grid links. In addition to the information services of the original smart grid, a variety of new services, including photovoltaic cloud network, energy storage cloud, vehicle network, new energy cloud, financial media cloud, comprehensive energy application platform, multi-station integration, data center, and source network charge interaction system, have been introduced, and new requirements have been put forward for the development of the existing power communication network, no matter from the service mode or the mode of networking [2]. In this context, the scientific information development technology model is taken as the standard to pay attention to the overall development and construction needs of the electric power industry, and the necessary information and system applications are implemented in combination with the actual monitoring model of the power Internet of Things network, and the standards of information exchange and communication protocols are defined.

Therefore, this paper studies the hybrid communication network protocol optimization technology of ubiquitous power Internet of Things to improve the efficiency and reliability of hybrid communication network.

## 2 Hybrid Communication Networking Protocol Optimization Technology for Ubiquitous Power Internet of Things

### 2.1 Extract Characteristics of Communication Networking Requirements

At present, the main business of ubiquitous power IOT can be divided into internal business and external business. Since there is little change in the mode of communication networking and carrying of ubiquitous power IOT, this paper focuses on the future external business needs of ubiquitous power IOT. In external services, this paper mainly considers three aspects of communication network bandwidth demand, delay demand and reliability demand, and refines them according to the collection demand and control demand [3]. In power IoT, bandwidth is used to describe the capacity of a link to transmit service data, the maximum amount of data that a particular link in a distribution network can carry in a unit time. The maximum data transfer rate generally satisfies the second theorem of Shannon, and the expression of the function is as follows:

$$u = d \cdot \log_2(1 + \alpha) \quad (1)$$

In Eq. (1),  $u$  represents the capacity of the channel used for data transmission;  $d$  represents the channel bandwidth;  $\alpha$  stands for signal-to-noise ratio. Ubiquitous power IOT services mainly focus on bandwidth requirements of more than 2 Mbit/s, with some system stations and clouds having different bandwidth requirements. According to the outline for the construction of ubiquitous power Internet of Things released by China Grid Corporation, the 2 Mbit/s business will be mainly considered, while for the interconnection of cloud platforms or other Internet of Things data platforms, the data traffic is usually convergent traffic, so the granularity is relatively high, and the data will depend on the number of terminals. Delay usually refers to the time spent from the

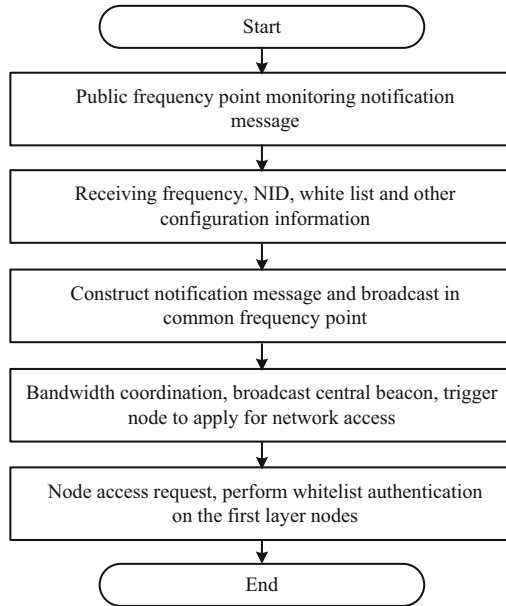
sending end to the receiving end of the power service packet, which is composed of transmission delay, propagation delay, processing delay and queuing delay. Transmission delay and propagation delay are considered in this paper. Transmission delay is composed of the time of collecting and encapsulating data packets, so it is related to the size of data packets, the way of encapsulation and the corresponding hardware. The propagation delay is related to the propagation of data in the communication link, so its size is determined by the propagation distance and the physical medium of the transmission link [4]. In the case of large datagram length, the transmission delay should be considered firstly, while in the case of small datagram length, the propagation delay is the main factor to be considered. Ubiquitous in the power Internet of Things, reliability can be measured by utilization to describe the current network link workload. Network utilization rate describes the weighted average utilization rate of all the links in a distribution communication network. When the network utilization is low, it means that the link has less traffic and does not produce large network delay. However, with the increase of network utilization, network traffic increases, resulting in the need for data in the queue waiting for the receiver to receive data in turn. At this point the network will significantly increase the latency. The relationship between latency and utilization can be expressed in formula:

$$L = \frac{L_0}{1 - \beta} \quad (2)$$

In Eq. (2),  $L$  represents the current network delay;  $L_0$  represents the delay of idle network state;  $\beta$  represents the current utilization of the network. According to the demand characteristics of ubiquitous power Internet of Things (IoT) communication networking, a hybrid networking architecture is established to meet the operational requirements of power grids and customers' demand for electricity.

## 2.2 Establish the Hybrid Communication Networking Architecture of Ubiquitous Power Internet of Things

According to the demand analysis of hybrid communication networking of ubiquitous power Internet of Things, combined with the deficiency of networking scheme of user electricity information collection, the networking architecture is designed from the perspective of dual-mode construction and with reference to the low-voltage power line broadband carrier communication standard. The networking process includes multi-frequency networking, whitelisting of incoming nodes and selection of proxy nodes. Firstly, multi-frequency network is designed to evenly distribute multiple networks in different working frequency points to reduce inter-network interference as much as possible. The new network can quickly grasp the use of frequency points. The overall networking process is shown in Fig. 1.



**Fig. 1.** Overall networking process

A whitelist is stored in the node that has access to the network, and the whitelist is authenticated after receiving the associated request message. In the process of networking, the node that has entered the network does not receive the associated request message for one consecutive beacon period and stops receiving. Whenever the whitelist changes, the network is reorganized. A few nodes in the whitelist change, reorganize the network can increase the network burden. In the hybrid communication network of the Ubiquitous Power Internet of Things, the whitelist is allocated to each node connected to the network, and changes in the whitelist do not necessitate the reconfiguration of the network. This improvement reflects the stability and reliability of the network and enables the network to quickly return to normal operation [5]. The configuration of the white list includes: clearing the white list (deleting the existing white list), replacing the white list (replacing the old white list with the new white list), adding the white list nodes (adding 1 to multiple nodes to the existing white list), deleting the white list nodes (deleting 1 to multiple nodes to the existing white list), and increasing or decreasing the number of white list nodes. Selecting the appropriate proxy node is the key to solve the problem of data transmission. The goal of agent selection is to minimize the number of agents and balance the load as much as possible. Proxy node selection mainly consists of two parts: node selection candidate agent and designated proxy node [6]. In this paper, the node selects candidate agents using three criteria: signal-to-noise ratio, hierarchy and communication success rate. SNR and communication success rate determine the link quality, and the level determines the number of data forwarding. The smaller the level is, the less the number of forwarding

is, and the higher transmission efficiency is. In a multi-standard decision, the formula for calculating the comprehensive weight is as follows:

$$q_i = \lambda_\alpha \cdot \alpha_i + \lambda_c \cdot c_i + \lambda_s \cdot s_i \quad (3)$$

In Eq. (3),  $q_i$  represents the weight of the monitored node  $i$ ;  $\alpha_i$  represents the link signal-to-noise ratio of the  $i$  node monitored;  $c_i$  represents the level of the node monitored;  $s_i$  represents the success rate of communication between node and node  $i$ . During networking, the success rate of communication is the number of beacon frames monitored.  $\lambda_\alpha, \lambda_c, \lambda_s$  correspond to weights of  $\alpha_i, c_i, s_i$ , respectively. The larger the weight  $q_i$  value is, the more suitable it is to be a candidate agent. When the agent node is specified, the candidate agent with the least number of child nodes is selected, and the hierarchy of the candidate agent meets the requirements, then it is designated as the agent node. The multi-standard decision method is used to achieve the goal of designing the hybrid communication network architecture with the minimum number of nodes and load balance.

### 2.3 Design the Routing Algorithm of Communication Network

For the hybrid communication network ubiquitous in the power internet of things, base station nodes need to manage, collect and maintain the information of all nodes in the network. Adopting appropriate communication group routing algorithm and organizing the nodes in the network according to certain topological structure and logical relationship can greatly facilitate the network management [7, 8].

Ant colony algorithm is used to select routing information in order to optimize the effect of the algorithm before designing the routing algorithm. In order to optimize the route and guarantee the communication efficiency, it is necessary to reset the obstacle avoidance rule. For the obstacle or interference that will not cause great influence on the communication signal, it is set to avoid the problem. When the communication route encounters this kind of obstacle, it is not necessary to replan the other route and reduce the distance of the communication route as far as possible. Multiple path nodes and obstacle nodes in the communication environment are represented by ant colony algorithm, and the optimal path of data transmission and the path after avoiding obstacle are expressed. Then, the ideal optimal path and the actual optimal path can be calculated to analyze the situation of avoiding different obstacles by ant colony algorithm:

$$q = \frac{f(u_x, v_x) - f(u_y, v_y)}{\varepsilon \sqrt{\sigma - 1}} \quad (4)$$

In Eq. (4),  $q$  refers to the obstacle identification parameters of ant colony,  $f$  refers to the identification function of different nodes,  $u, v$  refers to the identification results of coordinates of different nodes,  $\varepsilon$  refers to the number of key information identified, and  $\sigma$  refers to the total number of node information involved in path identification planning. The ability of the ant colony to recognize obstacles and routes in the environment can be obtained by the above formula. The recognition ability can be obtained by

adjusting the obstacle avoidance rules to enhance the pheromone selection ability of ant colony to critical path nodes and important obstacle nodes.

In this paper, the steps of designing a hybrid network routing algorithm for ubiquitous power Internet of Things are as follows: (1) Sending network broadcasts from the base station node to the surrounding, and initiating the networking process. Assuming that a total of  $A$  nodes receive netting packets from the base station node, the  $A$  node replies to the base station node in turn [9, 10]. The base station node receives a reply from the  $A$  node, indicating that the  $A$  node successfully joins the logical layer. (2) The  $A$  node that has joined the network shall be granted the authority of networking in turn, and shall send network broadcasting. Nodes that receive networked broadcasts may: A. be in the same logical layer as the sending node (logical layer 1); B. be in the next logical layer (i.e. have not joined the network); and C. be in the logical layer above. The specific processing process is as follows: if the receiving node and the sending node are located at the same layer, the receiving party shall record the corresponding receiving quality, and select two (at most two) nodes with the lowest receiving quality as its neighbor nodes, which shall be recorded in the local routing table. If the receiving node is in the next logical layer, the node sends a reply to the sender, joins the cluster with the sender node as the central node, and becomes a member of the cluster. After receiving the reply, the central node records the ID of the cluster-node in the local routing table, and reports the access information of the new node to the base station node [11]. If the receiving node is above the sending node, it is not processed. Node processing is completed with some delay until all nodes in Logical Layer 1 have sent their networked broadcasts. (3) Repeat step (2) until there is no receiving node at the next level (not yet joined the network). In the hybrid communication network of ubiquitous power Internet of Things, due to its time-varying topology, the channel quality may be degraded in the process of network operation, resulting in link failure and direct communication [12]. In order to ensure the reliability and stability of communication, it is necessary to adopt an effective link failure recovery mechanism by enabling the alternate path to restore communication. If an uplink failure occurs, the source node can select one of its neighbors as a temporary relay node to allow communication to proceed normally. If downlink failure occurs, the source node needs to add a failure flag to the original packet to notify it of downlink failure with the destination node. The node that receives the packet checks to see if it is the neighbor of the destination node [13, 14]. If so, the failure flag is removed and the packet is sent to the destination node. Conversely, the destination node returns the packet to the source node through the neighbor node after it receives the packet. At this point, all nodes have joined the network, ubiquitous IOT hybrid communication group network routing strategy is completed.

#### 2.4 Set up the Hybrid Networking Communication Protocol

Although the communication protocol based on PRIME standard has been well designed and has been applied in Europe and America, its popularity in our country is limited to a certain extent because it is based on the European power line communication industry standard, and its content and system are complicated, and the requirements for computing capacity and storage space of nodes are higher [15]. On the

basis of establishing hybrid communication network architecture and designing routing algorithm, the hybrid communication network protocol of ubiquitous power Internet of Things is optimized. Existing communication protocols are mainly used for local switching, without considering the coding mechanism of relay forwarding. Therefore, some fields such as relay routing table and hops are added to the data frame. This paper adopts CSMA/CA competition mechanism based on binary exponential backoff algorithm which is widely used in IEEE802.15.4 protocol to design the MAC layer protocol. The basic principle of the algorithm is: when the node has data to send, it first random backoff for a period of time, and then listen to the channel. If it is found that the channel is idle, it sends data; if it is found that the channel is busy, it will be randomly evaded for a period of time, repeat the previous step. If the number of retreats exceeds, the communication fails and the packet is dropped. The specific steps are as follows: (1) Initialize the parameters before sending the data by the node; (2) Randomly delay  $0 \sim (2^i-1)$  the evasion period ( $i$  is the minimum evasion index); (3) Detect whether the channel is idle, and if it is idle, send the data; otherwise, proceed to the next step; (4) Add 1 to the value of the evasion order, and set the minimum evasion index; and (5) Check whether the value of the evasion order is greater than the maximum evasion index, and if so, determine that the channel access fails and the packet is dropped; otherwise, go to step (2). The standard data frame is mainly composed of frame header, control code, source node, destination node, layer number, hop number, relay routing table, data length field, data field, check code and frame tail. Due to the strong attenuation, high noise and time-varying load impedance of power line, the length of data frame should not exceed 63 bytes in order to ensure the reliability of communication. The broadcast process is initiated by a base station node sending a data frame, and the node receiving the broadcast frame continues to broadcast the frame to its children until it traverses the network. The unicast process is initiated by a unicast data frame sent by a base station node to a specified destination node [16–18]. The base station node searches for the main route in the routing table of the base station node, and writes it into the relay routing table domain, and sets the number of hops as the number of node IDs. If a node finds itself in the first node address position in the relay routing table domain after receiving a data frame, it sets the source node to a MAC ID, deletes the first node address stored in the relay routing table domain, subtracts the number of path hops by 1, forms a new data frame, and sends it. This continues until the data frame reaches the destination node. After receiving the data frame, the destination node decides whether to reply to the base station according to the need. Based on the above process, this paper finishes the optimization design of hybrid communication network protocol for ubiquitous power Internet of Things.

### 3 Experimental Study

In order to test the application effect of the proposed hybrid communication networking protocol optimization technology for ubiquitous electric Internet of Things, a simulation experiment was designed to verify the performance of the communication protocol.

### 3.1 Experimental Preparation

This experiment mainly uses MATLAB TrueTime network simulation toolbox to carry out simulation experiment verification and result analysis. In order to compare the routing performance of different communication protocols, the same node distribution and physical layer parameters are used in the simulation process. Set in the 100 m100 m network, there are a total of 15 randomly distributed nodes, among which node 1 is the base station node. The transmission rate is 128.6 kbps. The location distribution of each node in the network is shown in Fig. 2.

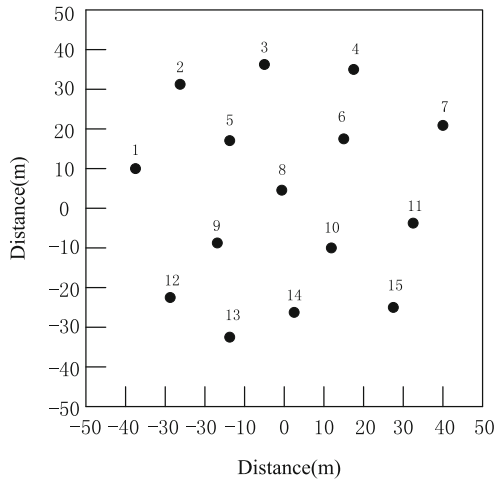
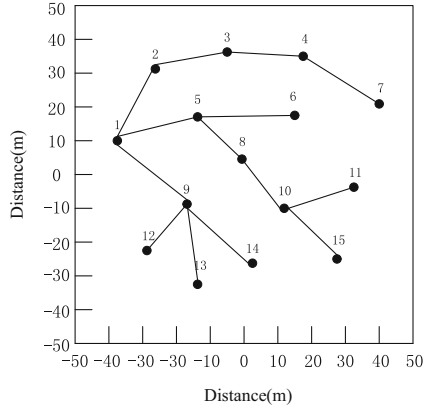


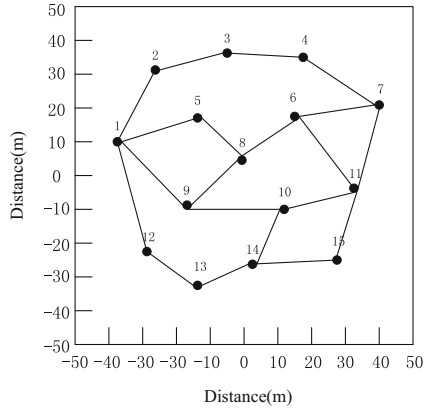
Fig. 2. Distribution of nodes

### 3.2 Communication Protocol Routing Overhead Test

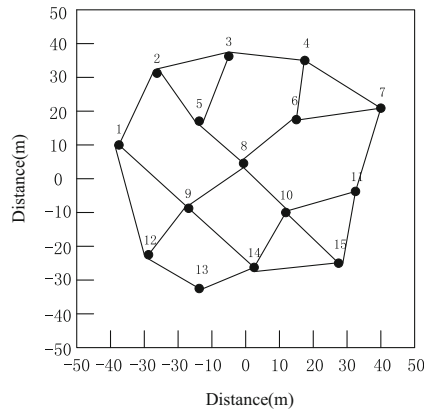
In this experiment, the optimized technology communication protocol proposed in this paper was set as the experimental group, and the networking communication protocol based on PRIME standard and the improved PRIME communication protocol were selected as the control group to conduct a comparative experiment. The three communication protocols were used to form the network respectively, and the logical topology of the network nodes finally formed was shown in Fig. 3.



(a) Prime Agreement



(b) Improvement of the PRIME agreement



(c) This paper designs a communication protocol

Fig. 3. Communication protocol networking structure

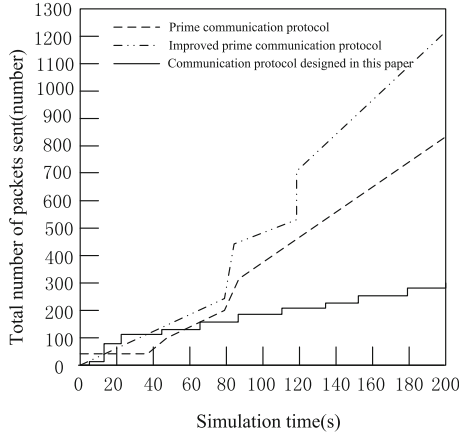


As can be seen from Fig. 3, the network structure composed of PRIME communication protocol is a tree type, while the improved PRIME communication protocol and the communication protocol designed in this paper form a mesh topology structure with chordal connection. In order to comprehensively compare the performance of the above communication protocols, the routing cost is selected as the evaluation index. Routing overhead refers to the number of all control packets transmitted in the network during networking and maintenance. It reflects the effective transmission efficiency of the network. The routing overhead of each node of the three communication protocols is compared, and the comparison results are shown in Table 1.

**Table 1.** Comparison results of routing overhead of each node

The serial number	Number of packets sent (PCS)		
	PRIME communication protocol	Improve PRIME communication protocol	This paper designs a communication protocol
1	578	235	46
2	574	227	48
3	562	236	45
4	549	242	46
5	563	247	48
6	572	235	47
7	568	238	43
8	570	241	46
9	564	228	47
10	575	236	48
11	572	245	49
12	569	246	43
13	543	241	44
14	572	239	46
15	568	237	48

According to the comparison results of routing overhead among nodes in Table 1, the communication overhead of each node of PRIME communication protocol is 567 data packets, the communication overhead of each node of the improved PRIME communication protocol is 238 data packets, and the communication overhead of each node of the communication protocol designed in this paper is 46 data packets. The total routing overhead of the three communication protocols was further compared, and the simulation time was 200 s. The test results were shown in Fig. 4.



**Fig. 4.** Test results of total routing cost of communication protocol

According to the test results of total routing cost in Fig. 4, under the condition of simulation time of 200 s, the total routing cost of the communication protocol designed in this paper is 284 packets, which is 920 and 548 packets less than the total routing cost of PRIME communication protocol and improved PRIME communication protocol respectively. The reasons for the above results are as follows: the registration, upgrade and maintenance process of PRIME communication protocol node is complex, which involves the exchange of a large number of data packets, making the total routing overhead high. And the frequency of beacons sent by base station nodes is higher than that of other nodes, which increases the routing overhead of each node. The improved PRIME communication protocol simplifies the node upgrade process and stipulates that the beacon transmission frequency of the base station node and other nodes is the same, thus reducing the total routing overhead and the routing overhead of each node. In the communication protocol designed in this paper, the network entry node has the authority of independent networking, and the child node can join the network directly through the parent node without the permission of the base station node, which greatly simplifies the networking process. The network maintenance mechanism was also removed. Therefore, compared with the previous two, the routing overhead is greatly reduced.

## 4 Conclusion

This paper studies the optimization technology of hybrid communication networking protocol for ubiquitous power Internet of Things, analyzes and proposes the optimized hybrid networking architecture, and proposes the appropriate communication protocol based on the hybrid networking architecture. The experimental results show that compared with the existing communication protocols, the communication protocol designed in this paper has better performance, can greatly reduce the routing overhead,

and has a higher practical value. The optimization of the hybrid communication network protocol can effectively improve the data transmission efficiency of the power communication network, enhance the application effect of the power Internet of Things, and improve the reliable guarantee for the operation efficiency of the power industry. If the hybrid communication network protocol optimization method proposed in this paper is applied to the medical Internet of Things, it can also effectively improve the efficiency of medical information transmission and promote the development of medical digitalization.

However, due to my limited ability, research time and experimental environment conditions, further improvement is needed in future research work. All the research work in this paper relies on the simulation platform to carry out quantitative demonstration on the reliability of the communication network, and lacks specific field experimental data for comparison. Therefore, in future research work, it is necessary to fully compare the test results of field experiments to make the simulation results closer to the actual situation.

## References

1. Zhang, N., Yang, J., Wang, Y., et al.: 5G communication for the ubiquitous Internet of Things in electricity: technical principles and typical applications. In: Proceedings of the CSEE, vol. 39, no. 14, pp. 4015–4024 (2019)
2. Cui, J., Wang, S., Xin, Y., et al.: Research on data management mechanism of ubiquitous power Internet of Things from the consortium blockchain perspective. *Inf. Sci.* **38**(4), 37–43 (2020)
3. Fu, Z., Li, X., Yuan, Y.: Research on key technologies of ubiquitous power Internet of Things. *Electr. Power Constr.* **40**(5), 1–12 (2019)
4. Wang, H., Tang, R., Wu, Y., et al.: Research and application of smart meter reading technology based on HPLC in customer side universal power Internet of Things. *Power Syst. Prot. Control* **48**(3), 92–98 (2020)
5. Zheng, X., Ying, Z., Wang, Q., et al.: Research on wireless communication access technology for ubiquitous power Internet of Things. *Electr. Power Constr.* **40**(11), 16–23 (2019)
6. Zhao, C., Qiu, X., Zhao, Y., et al.: A new UPFC system in the framework of ubiquitous power Internet of Things. *Electr. Power Constr.* **41**(1), 39–44 (2020)
7. Hou, X., Liu, X., Zheng, K., et al.: Technical prospect of a new generation smart meter in the ubiquitous power Internet of Things environment. *Electr. Meas. Instrum.* **57**(9), 128–131 (2020)
8. Xiao, Z., Xin, P., Liu, Z., et al.: An overview of planning technology for active distribution network under the situation of ubiquitous power Internet of Things. *Power Syst. Prot. Control* **48**(3), 43–48 (2020)
9. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. *Mob. Netw. Appl.* **26**(5), 1891–1903 (2021). <https://doi.org/10.1007/s11036-020-01725-x>
10. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. *Multimedia Tools Appl.* **76**(4), 5787–5802 (2014). <https://doi.org/10.1007/s11042-014-2408-1>

11. Liu, S., Liu, G., Zhou, H.: A Robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–17 (2018). <https://doi.org/10.1007/s11036-018-1134-8>
12. Xi, L., Yu, L., Zhang, X., et al.: Automatic generation control of ubiquitous power Internet of Things integrated energy system based on deep reinforcement learning. *SCIENTIA SINICA Technologica* **50**(2), 221–234 (2020)
13. Liu, Y., Pan, J., Wang, W., et al.: Research on intelligent sensing terminal technology of a distribution transformer for ubiquitous power Internet of Things. *Power Syst. Prot. Control* **48**(16), 140–146 (2020)
14. Shukla, V., Chaturvedi, A., Srivastava, N.: A secure stop and wait communication protocol for disturbed networks. *Wirel. Pers. Commun.* **110**(2), 861–872 (2019). <https://doi.org/10.1007/s11277-019-06760-w>
15. Matheus, L.M., Vieira, A.B., Vieira, M., et al.: DYRP-VLC: a dynamic routing protocol for wireless ad-hoc visible light communication networks. *Ad Hoc Netw.* **94**(15), 101941 (2019)
16. Delai, A.L., Miyadaira, A.N., Lima, T.C.: AMASP (ASCII master slave protocol): a lightweight MODBUS based customizable communication protocol for general applications. *J. Comput. Inf. Syst.* **34**(1), 12–23 (2019)
17. Ünlü, B., Özceylan, B., Baykal, B.: IPBM: an energy efficient reliable interference-aware periodic broadcast messaging protocol for MANETs. *Wirel. Netw.* **25**(5), 2769–2787 (2019). <https://doi.org/10.1007/s11276-019-01992-x>
18. Haq, A., Faheem, Y.: A peer-to-peer communication based content distribution protocol for incentive-aware delay tolerant networks. *Wireless Netw.* **26**(1), 583–601 (2019). <https://doi.org/10.1007/s11276-019-02167-4>



# Forecasting Method of Monthly Clearing Price Under the Background of Continuous Adjustment of Power Market Supply and Demand

Guo-bin Wang<sup>1(✉)</sup>, Wen-tao Xu<sup>1</sup>, Le-le Wang<sup>1</sup>, Jing An<sup>1</sup>,  
and Yang Bai<sup>2</sup>

<sup>1</sup> Marketing Service Center (State Grid Ningxia Electric Power Co., Ltd. Metering Center), State Grid Ningxia Electric Power Co., Ltd., Ningxia 750000, China  
gslwxyx2021@163.com

<sup>2</sup> State Grid (Ningxia) Integrated Energy Service Co., Ltd., Ningxia 750000, China

**Abstract.** The monthly clearing price forecasting method is commonly used. The model input is used to construct a monthly electricity clearing price data set. The collected electricity monthly clearing price time series characteristics are too single, resulting in a large average absolute percentage error of the monthly clearing price forecast. For this reason Propose a method for forecasting monthly clearing prices in the context of the continuous adjustment of supply and demand in the electricity market. To study the impact of the continuous adjustment of power market supply and demand on the monthly clearing price, design the electricity monthly clearing price forecasting process, and use the normalization method to preprocess the data; extract time information and load information to construct a training sample set for the monthly clearing price of electricity; The non-linear mapping relationship between electricity price and various influencing factors, using BP neural network, establishes a monthly electricity clearing price prediction model, and predicts the monthly electricity clearing price. The experimental results show that the average absolute percentage error of the research method predicting the monthly electricity clearing price is smaller than the two commonly used methods, and it has better prediction accuracy of the monthly electricity clearing price.

**Keywords:** Electricity market · Market supply and demand · Adjustment of supply and demand · Monthly clearing · Price forecast

## 1 Introduction

The traditional system of the power industry is a vertical monopoly integrating generation, transmission, distribution, and supply. This system played an indispensable role in the initial stage of power development. It not only improved the efficiency of regional grid planning, but also benefited the internal power system. Coordination and cooperation. With the rapid development of the socialist economy, various problems in

this system have become increasingly prominent, such as the slow development of power companies, bloated personnel and organizations, high electricity prices, and poor services provided [1]. However, with the further deepening of the restructuring and deregulation of the power industry worldwide, the reform of the power market has become an inevitable trend in the development of the power industry. Establishing a reasonable electricity price mechanism is the core issue of the electricity market reform. From a long-term perspective, a transparent electricity price mechanism is more conducive to attracting power investment, thereby ensuring the stable and sustainable development of the power industry.

My country's power industry strongly supports socialist economic construction. However, the monopoly of power generation, transmission, and distribution has increasingly restricted the development of my country's power industry, and conflicts with the construction of the socialist market economic system have become increasingly prominent, resulting in unfair social distribution problems. Increasingly prominent [2]. On the other hand, theoretically speaking, electric energy is a commodity, which has value and use value in itself, and this commodity should enter the market competition. The monopolistic operation under the planned economy is the direct cause of the non-separation of government and enterprise. Too much government intervention makes power companies have no independent market players and their development has no vitality. Based on this, my country has clearly put forward the policy of "separating power plants and grids, and bidding for the Internet", which intends to introduce competitive factors into the power industry, enhance the market competitiveness of power companies, and establish a standardized and orderly power market. As an independent economic entity in the market environment, power generation companies no longer arrange power generation based on planned electricity in the traditional way, but participate in market competition to fight for power generation indicators and draw on-grid electricity prices based on market conditions. Accurate electricity price forecasting can provide reliable reference information for the strategic quotation of power generation companies to maximize revenue; similarly, electric energy users also need accurate electricity price information to guide the signing of long-term bilateral contracts and make the best purchase strategy.

Therefore, the more accurate the short-term electricity price forecast is provided, the more reliable information market participants can obtain in the transaction, which provides a favorable reference for them to formulate the optimal quotation strategy, and thus obtains greater benefits. However, the reform of my country's electricity market is still in its infancy. The market structure and trading rules are still changing. The market system is not perfect. Electricity prices tend to fluctuate and jump. It is very difficult to accurately predict short-term electricity prices [3]. Therefore, short-term electricity price forecasting has become one of the urgent problems to be solved in the electricity market.

In the electricity market environment, electricity price is a core indicator to evaluate the efficiency of market competition, reflecting the operation status of the electricity market. As an important part of the electricity market bidding system, short-term electricity price forecasting in the electricity market can provide effective information for all participants in the electricity market, which is of great significance for maximizing corporate profits.

At present, short-term electricity price forecasting methods mainly include time series forecasting, artificial neural network method, wavelet theory, gray model method, combined model method and other methods [4]. Reference [5] tries to apply it to the electricity market for the first time, and takes into account different power price influence factors, and the accuracy of short-term electricity price forecasts is gradually improved. Reference [6] first uses the fuzzy C-means clustering method to classify market information, and then uses neural networks to make predictions for each category, and the prediction error is low.

In order to solve the shortcomings of traditional methods, this paper designs a new monthly clearing price forecasting method under the background of constant adjustment of supply and demand in the electricity market. In this method, the monthly settlement price prediction process is designed according to the influence of constant adjustment of supply and demand on the monthly settlement price. Then the normalized method is used to preprocess the data and extract the time information and load information, so as to construct the training sample set of monthly settlement electricity price. On this basis, using the nonlinear mapping relationship between the price of electricity and the influencing factors, the BP neural network is used to establish the monthly clear price prediction model.

## **2 Forecasting Method of Monthly Clearing Price Under the Background of Continuous Adjustment of Power Market Supply and Demand**

### **2.1 Study the Impact of the Continuous Adjustment of Supply and Demand in the Electricity Market on the Monthly Clearing Price**

In the context of the continuous adjustment of supply and demand in the electricity market, the Liku model and the bilateral transaction model are currently the two relatively mature electricity market transaction models [7]. Electricity prices are also affected by factors such as social development, government activities and policies, environmental protection, power technology, and supply and demand.

Market factors. It mainly includes the entire historical load of the power market, the load rate of the system, overcapacity or/shortage, the historical reserve of energy, hydroelectric power generation, power generation capacity, power system constraints, nuclear power generation and thermal power, etc. For example, when the system load increases, the electricity price will increase accordingly.

Non-strategic uncertainties include weather, temperature, crude oil prices, natural gas prices, fuel prices, energy reserves, load forecasts, and dew point temperature. For example, electricity needs depend on the environment, especially the daily temperature, and changes in weather affect the load and price. In addition, if there is not enough energy reserves, consumers will face a lack of energy, which will cause a balance between supply and demand and cause price changes.

Random uncertainties mainly include production outages, line emergency, line blockage, etc. For example, some power plants are far away from consumers, and

transmission requires corresponding network facilities. There will be some physical factors that hinder the energy trading of market participants.

The behavior index mainly refers to the historical electricity price, elastic load, bidding strategy, and electricity price shock index.

Time effects include settlement period, day, month, season, holiday, summer index and winter index. For example, due to weather in summer and winter, load demand will increase, and electricity price fluctuations will increase accordingly. Electricity prices will also increase during a certain peak period of the working day, and electricity prices will also fall when electricity consumption is less at midnight. Therefore, compared with general commodities, the setting of electricity prices is more complicated.

At present, the domestic and foreign electricity price pricing system is mainly divided into three categories: non-market pricing, market pricing (competitive pricing), and contract pricing (agreement pricing). However, in the context of the continuous adjustment of supply and demand in the electricity market, competition is usually introduced in the on-grid electricity price, and electricity prices are set through the market; the government regulates transmission and distribution prices and sets prices independently; the government regulates the sale of electricity prices based on market leadership. In different markets, due to different operating models and different bidding mechanisms, different types of price bidding have appeared.

For the first mock exam, there is a system marginal price bidding and bidding system bidding. Under the bilateral transaction, there are market clearing price and high-low matching transaction price. When a power company purchases electricity from a power producer through bidding as a monopoly, it then sells electricity to consumers through bidding or direct pricing as a monopoly. This price is called systematic marginal price bidding; The price based payment bidding is the actual price settlement method adopted by the power trading center, according to the market demand to the power suppliers' respective quotation; High matching transaction price is similar to the transaction mechanism of stock price, that is, the transaction is matched in the power trading center according to the principle of high and low matching of purchase price and selling price.

Based on the above analysis, there are five main factors that cause price fluctuations:

1. Electricity demand, an important factor of the spot price is the total demand of the system.
2. It is the weather conditions, the power demand depends on the environmental conditions, especially the daily temperature. Weather fluctuations will affect demand, so spot prices will also be affected.
3. Fuel price. Fuel cost is one of the main parts of power generation cost, and its changes have a significant impact on the spot price of electricity.
4. Available transmission capacity. Electricity is generally provided by generators far away from consumers. It is sent to consumers through the transmission network facility. If there are some physical constraints in the transmission network, it will become an obstacle for market participants to buy or sell energy. This will affect the changes in spot prices.



5. Energy reserve. Adequate energy reserve is an important factor to ensure the spot price of electricity. If there is a sudden increase in power demand, and if there is enough power reserve capacity to use, then the service to consumers will be guaranteed. But if there are not enough available reserves, consumers will face a lack of energy, which will increase the balance of power spot prices between supply and demand.

## 2.2 Price Prediction Process and Data Preprocessing

### Forecast Form

Based on the research results in the previous section that the monthly clearing price is affected by the continuous adjustment of power market supply and demand, the formal calculation process of the designed monthly clearing price forecast is as follows:

If given a set  $(y_1, y_2, \dots, y_t)$  and  $x_{t+1}$ , estimate  $Y_{t+1}$ , the goal is to minimize  $E(Y_{t+1}, y_{t+1})$ . Among them,  $(y_1, y_2, \dots, y_t)$  is the observed value of the target variable  $y$  at time  $t$ , and  $t = 1, 2, \dots$ ;  $x_{t+1}$  is the observed value of the input vector  $x$  at time  $t + 1$ , the input vector may include some observed values of the target vector  $y$ ;  $Y_{t+1}$  is the predicted value at time  $t + 1$ ,  $y_{t+1}$  is the actual value of  $t + 1$  at time;  $E$  is the error function, used to measure how close the predicted value is to the actual value. The task of prediction is to estimate the relationship between the estimated input vector and the predicted value:

$$Y = f(x, a) \quad (1)$$

In the formula (1),  $Y$  represents the prediction model;  $a$  represents the parameters of the model [8].

According to the above-mentioned monthly clearing price forecast form, the monthly clearing price is predicted. There will be a certain gap between the predicted value of the monthly clearing price and the true value, that is, an error. Therefore, it is necessary to evaluate the forecast value to reduce the forecast error of the monthly clearing price.

In error evaluation, the larger the error, the smaller the accuracy of the prediction; on the contrary, the smaller the error, the higher the accuracy of the prediction.

Generally, there are two reasons for errors. One is sample error, which is mainly caused by sample inaccuracy; the second type of error is called model error, and different prediction models produce different results. The error is also different. At present, in the field of monthly clearing price forecasting, in order to evaluate the efficiency of the monthly clearing price forecasting model, error criteria such as absolute error  $M$ , mean square error  $S$ , average absolute percentage error  $P$ , root mean square error  $R$  are often used as judgments Index, its calculation formula is as follows:

$$\begin{aligned}
M &= \frac{1}{n} \sum_{i=1}^n |Y_i(t) - y_i(t)| \\
P &= \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i(t) - y_i(t)}{Y_i(t)} \right| \\
R &= \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i(t) - y_i(t))^2} \\
S &= \frac{1}{n} \sum_{i=1}^n (Y_i(t) - y_i(t))^2
\end{aligned} \tag{2}$$

In the formula (2),  $n$  represents the predicted number of points;  $Y_i$  represents the true value;  $y_i$  represents the predicted value. In the monthly clearing price data, a large number of peak monthly clearing prices and abnormal monthly clearing prices directly affect the distribution of monthly clearing prices. Sometimes the value of the monthly clearing price may be close to zero, which will make the average absolute percentage error very small. Especially when the monthly clearing price has a negative value, if the positive and negative values are offset, the error metric will be small, and the result of this judgment is not very reasonable. Therefore, it is not appropriate to use the traditional definition of the average absolute percentage error to analyze the prediction results. For this reason, the weighted average absolute error  $D$  and Theil index  $U$  are used to evaluate the performance of the prediction method. The calculation formula is as follows:

$$\begin{aligned}
D &= \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - y_i|}{Y} \\
0 < U &= \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i(t) - y_i(t))^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n Y_i(t)^2 + \frac{1}{n} \sum_{i=1}^n y_i(t)^2}} < 1
\end{aligned} \tag{3}$$

Based on the above content, the designed monthly clearing price forecast form and the evaluation method of forecast error, the designed monthly clearing price forecast steps are as follows:

1. Select input data. Usually comes from historical electricity price data in a certain market.
2. Data preprocessing. In order to obtain accurate predictions, data needs to be pre-processed, such as data cleaning, missing value supplementation, outlier removal, data feature extraction, and so on.
3. Choose the right model. After a simple statistical analysis of the input data, such as mean and volatility, a suitable prediction model is selected based on some implicit information. The important factors for choosing a suitable model also include the scope of the prediction and the accuracy of the prediction.

4. Optimize the parameters of the model and check the predictive efficiency of the model.
5. Model evaluation. Some error criteria are used to evaluate the degree of fit between the predicted value of the model and the true value, and to measure the prediction accuracy of the model.
6. Successfully apply the model to the actual electricity price forecast.

### **Preprocessing of Electricity Monthly Clearing Price Data**

Changes in electricity prices are affected by various factors, so the time series of electricity prices are almost always non-stationary and non-linear. If some data preprocessing techniques are used before prediction to reduce some data fluctuations, and then a prediction model is constructed on the processed data, the final prediction accuracy is often higher than that of direct prediction. At present, the data preprocessing technology used in the field of electricity price forecasting mainly includes normalization technology.

1. Data normalization. This technology is a data processing method that is often used at present. If the magnitude of the data in each dimension differs greatly, it will lead to a large prediction error. Therefore, data normalization can eliminate the difference between input data and output data. At the same time, in order to speed up the training speed of the neural network, data normalization is also adopted. At present, there are mainly two commonly used data normalization methods. The first is the maximum minimum method. The specific form is as follows:

$$x_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

In the formula (4),  $x_k$  is the original data sequence;  $x_{\max}$  is the maximum value in the data sequence;  $x_{\min}$  is the minimum value in the data sequence [9].

Based on the price prediction process and data preprocessing technology designed this time, in the context of the continuous adjustment of supply and demand in the electricity market, the BP neural network is used to predict the monthly clearing price of electricity.

## **2.3 Constructing a Training Sample Set of Electricity Monthly Clearing Prices**

### **Time Information Extraction**

The monthly electricity clearance data has significant periodicity, but this periodicity contains very important information. For this reason, the time characteristics of the monthly electricity clearance data are extracted according to Table 1.

**Table 1.** Time characteristics of monthly electricity clearance data

Serial number	Feature name	Description
1	Day of the week	Day of the week
2	Day of the year	The day of the year
3	Day of the month	The day of the month
4	Week of the year	Week of the year
5	Hour of the day	Which hour of the day
6	Month of the year	Which months of the year

Electricity prices are obviously cyclical, and this cyclicity is determined by a number of aspects. For example, the price of electricity on day  $d$  and hour  $\tau$  is definitely the same as the price of electricity on day  $d - 1$  and hour  $\tau$ . There is a certain periodicity, etc., so based on the above questions, six time features related to electricity price forecasts are extracted.

### Load Information Extraction

Regional load and overall load are two very important features provided by the data set. Because the price of electricity fluctuates, the electric load plays a very important role in it, because if the electric load is very high, in order to reduce The pressure of equipment, the market will increase the price of electricity, which will cause some price-sensitive users to reduce the generation of unnecessary electricity, thereby reducing the overall electricity consumption, and this feature is lagging, that is, at time  $\tau$  It is found that after the power load increases, the power price will be increased at time  $\tau + 1$  to reduce the power consumption, and there is a great correlation between the power price and the load information, and the impact of the load characteristics on the power price is fully considered, so the following table is extracted. Table 2 shows the electrical load characteristics.

**Table 2.** Electrical load characteristics

Serial number	Feature name	Description
1	t value from $N \times 24$ h earlier	$N \times$ the overall load value 24 h ago
2	z value from $N \times 24$ h earlier	$N \times$ the overall load value 24 h ago
3	t value from $N$ h earlier	Overall load value $N$ hours ago
4	z value from $N$ h earlier	Area load value $N$ hours ago
5	Difference between total load and zonal load	Difference between regional load and overall load
6	Difference between zonal load and $N \times 24$ h earlier zonal load	The difference between the area load and the area load $N \times 24$ h ago
7	Difference between total load and $N \times 24$ h earlier total load	The difference between the overall load and the overall load $N \times 24$ h ago

As shown in Table 1, where N is the dimension of the corresponding feature. Take  $N \times$  the overall load 24 h ago as an example. If N is 1, the feature is a one-dimensional and only contains the overall load 24 h ago. If N is 3, the feature includes three dimensions, namely, the overall load at 24 h, 48 h, and 72 h ago.

### 2.4 Establish a Model for Forecasting Electricity Monthly Clearing Prices

Since the electricity price model of the power system reflects the non-linear mapping relationship between the electricity price and various influencing factors, the BP neural network is a multilayer non-linear mapping network that uses the minimum mean square error learning method to minimize its evaluation function., Complete the mapping of input signal to output mode [10]. It can realize complex and highly non-linear mapping for complex type pattern recognition. To this end, the BP neural network is used to predict the monthly clearing price of electricity under the background of the continuous adjustment of supply and demand in the electricity market.

BP neural network is a three-layer network structure, which is composed of input layer, hidden layer and output layer. Each layer may contain a different number of neurons. At different layers, each neuron has a nonlinear transfer function. The basic structure of BP neural network is shown in Fig. 1.

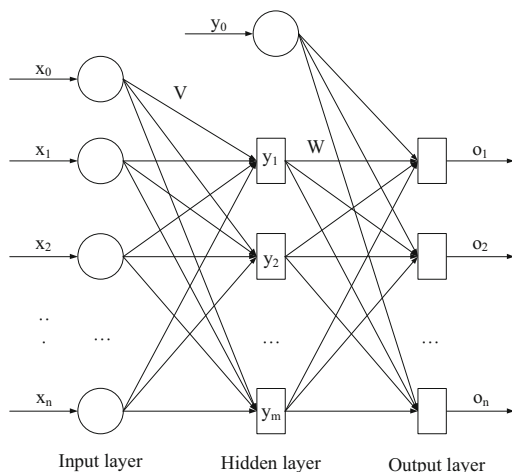


Fig. 1. Basic prediction structure of BP neural network

Figure 1, the input vector of BP neural network is  $X = (x_1, x_2, \dots, x_i, \dots, x_n)^T$ ,  $x_0 = -1$  is set for the hidden layer neuron reference threshold; the implicit output vector is  $Y = (y_1, y_2, \dots, y_j, \dots, y_m)^T$ ,  $y_0 = -1$  is the output layer neuron reference threshold and Set; the output vector of the output layer is  $O = (o_1, o_2, \dots, o_k, \dots, o_l)^T$ ; the expected output vector is  $C = (c_1, c_2, \dots, c_k, \dots, c_l)^T$ . The weight matrix between

the input layer and the hidden layer is represented by  $V$ :  $V = (V_1, V_2, \dots, V_j, \dots, V_m)^T$ , where the column vector  $V_j$  is the weight vector corresponding to the  $j$  neuron in the hidden layer; the weight matrix between the hidden layer and the output layer is represented by  $W$ :  $W = (W_1, W_2, \dots, W_k, \dots, W_l)^T$ , The column vector  $W_k$  is the weight vector corresponding to the  $k$  neuron in the hidden layer.

Based on the basic prediction structure of the BP neural network selected in this experiment, the unipolar Sigmoid function is selected as the transfer function  $f(x)$  of the prediction model:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

The basic prediction structure of the BP neural network selected this time, the input layer calculation formula is:

$$\begin{aligned} o_k &= f(\text{net}_k) \quad k = 1, 2, \dots, l \\ \text{net}_k &= \sum_{j=0}^m w_{jk} y_j \quad k = 1, 2, \dots, l \end{aligned} \quad (6)$$

In the formula (6),  $\text{net}$  represents the number of input networks in the BP neural network.

The calculation formula for the output layer is:

$$\begin{aligned} y_i &= f(\text{net}_j) \quad j = 1, 2, \dots, m \\ \text{net}_j &= \sum_{i=0}^n v_{ij} x_i \quad j = 1, 2, \dots, m \end{aligned} \quad (7)$$

In addition to the above calculation process of the input layer and output layer, it is also necessary to determine the learning algorithm of the network. In this study, considering the training error existing in the prediction process, two processes of forward propagation and reverse broadcasting of errors are adopted. As the learning algorithm of this study, the specific process is as follows: when forward propagation, the input samples are transferred from the input layer The input is processed by various layers and then passed to the output layer. If the actual output of the output layer does not match the expected output, then it goes to the error back propagation stage. Error backpropagation is to pass the output error back to the input layer layer by layer through the hidden layer in some form, and apportion the error to all the units of each layer, so as to obtain the error signal of each layer unit, and this error signal is used as a correction for each unit The basis of the weight [11]. The weight adjustment process of each layer of this signal forward propagation and error backward propagation is carried out in a round-robin fashion. The process of constant weight adjustment is the learning and training process of the network. This process continues until the network output error is reduced to an acceptable level, or until the preset number of learning times. Then the network error definition and weight adjustment process are as follows:

When the network output is not equal to the expected output, there is an output error  $e$ :

$$e = \frac{1}{2}(c - o)^2 = \frac{1}{2} \sum_{k=1}^l (c_k - o_k)^2 \tag{8}$$

Expanding the output error shown in formula (1) to the hidden layer and the input layer, there are:

$$e = \frac{1}{2} \sum_{k=1}^l [c_k - f(\text{net}_k)]^2 = \frac{1}{2} \sum_{k=1}^l \left[ c_k - f \left( \sum_{j=0}^m w_{jk} y_j \right) \right]^2 \tag{9}$$

It can be seen from formula (9) that the network error is a function of the weights  $w_{jk}$  and  $v_{ij}$  of each layer, so adjusting the weights can change the error  $e$ . Obviously, the principle of adjusting the weight is to continuously reduce the error, so the adjustment of the weight should be proportional to the gradient of the error, that is:

$$\begin{aligned} \Delta w_{jk} &= -\eta \frac{\partial e}{\partial w_{jk}} \quad j = 1, 2, \dots, m; \quad k = 1, 2, \dots, l \\ \Delta v_{ij} &= -\eta \frac{\partial e}{\partial v_{ij}} \quad j = 1, 2, \dots, m; \quad k = 1, 2, \dots, l \end{aligned} \tag{10}$$

In the formula (10), the negative sign represents the gradient descent, and the constant  $\eta \in (0, 1)$  represents the proportional coefficient, which reflects the efficiency of learning in training [12]. Under the effect of this learning algorithm, the flow characteristics of the predicted model signal constructed this time are shown in Fig. 2.

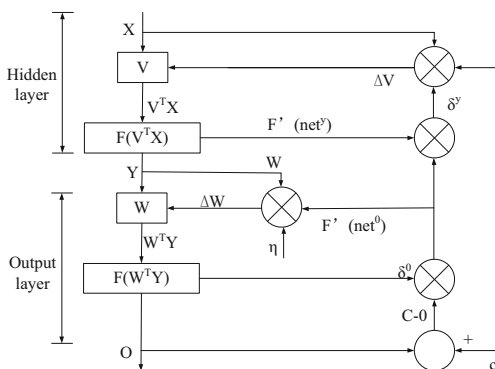


Fig. 2. The flow of predictive model signals

It can be seen from Fig. 2 that the forward process: After the input signal  $X$  enters from the input layer, the output signal  $Y$  of the layer is obtained through the internal

weight vector  $V_j$  of each node of the hidden layer; the signal is input forward to the output layer, and the weight of each node is Vector  $W_k$ , get the output  $O$  of this layer. The reverse process is: the expected output  $c$  in the output layer is compared with the actual output  $O$  to obtain the error signal  $\delta^0$ , from which the adjustment of the weight of the output layer can be calculated; the error signal  $\delta^0$  is transmitted back to the hidden layer through the vector of each node of the hidden layer each node obtains the error signal  $\delta^y$  of the hidden layer, from which the adjustment amount of the weight of the hidden layer can be calculated [13].

In summary, the monthly electricity clearing price forecast model established this time, the process of predicting the monthly electricity clearing price is as follows:

1. Initialization: Assign random numbers to the weight matrix  $W$  and  $V$ , set the sample mode counter  $p$  and the training times counter  $q$  to 1, the error  $e$  to 0, the learning rate  $\eta$  to a decimal in the interval (0,1), The accuracy  $e_{\min}$  achieved after training is set to a positive decimal.
2. Input the training samples and calculate the output of each layer: use the current samples  $X^p$  and  $c^p$  to assign values to the vector arrays  $X$  and  $c$ , and calculate the components of  $Y$  and  $O$ .
3. Calculate the output error of the network: suppose there are  $P$  pairs of training samples, the network has different errors  $e^p = \sqrt{\sum_{k=1}^l (c_k^p - o_k^p)^2}$  for different samples, and the root mean square error  $e_R = \sqrt{\frac{1}{P} \sum_{p=1}^P (e^p)^2}$  is used as the total error of the network.
4. Calculate the error signal of each layer: calculate  $\delta_k^0$  and  $\delta_j^k$ .
5. Adjust the weights of each layer: Calculate the components of  $W$  and  $V$ .
6. Check whether one round of training is completed for all samples: if  $e_R < e_{\min}$ , the training is over, otherwise  $e = 0$  and  $p$  are set to 1, and return to step 2.
7. After training, take the quantity to be demanded as input, and the output obtained is the quantity to be demanded.

### 3 Experiment and Analysis

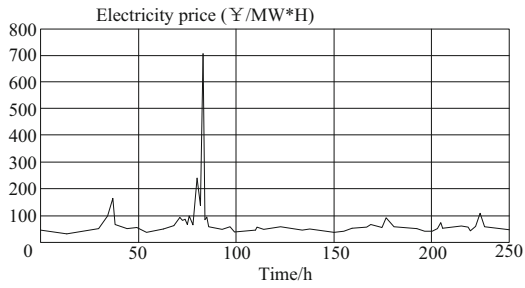
Two sets of commonly used forecasting methods are selected, and the monthly clearing prices of power plants in a certain area are selected as the data set of this experiment by means of comparative experiments, and the three sets of forecasting methods are compared to predict the average absolute percentage error of the monthly clearing prices.

#### 3.1 Experiment Preparation

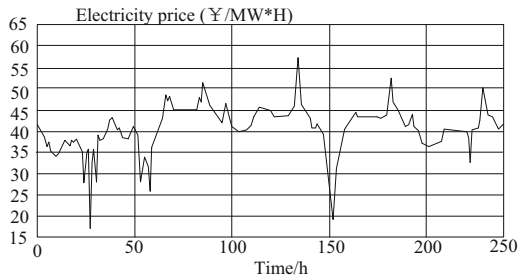
In this experiment, the monthly clearing price of power plants in a certain region is selected. According to the climate of the region, the most typical month is selected from the four seasons of summer, autumn, spring and winter as the three sets of prediction data, as shown in Fig. 3. As shown, and use one week's historical data to predict the electricity price on the last day of the week.



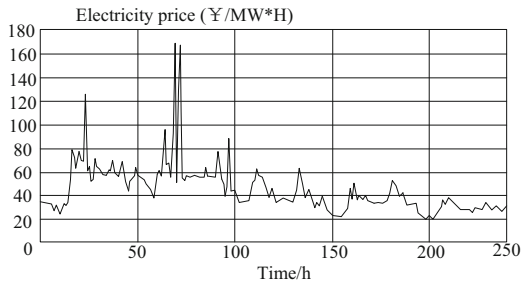
It can be seen from Fig. 3 that the characteristics of electricity price fluctuations in these four months are different. The price curve is constantly changing at different



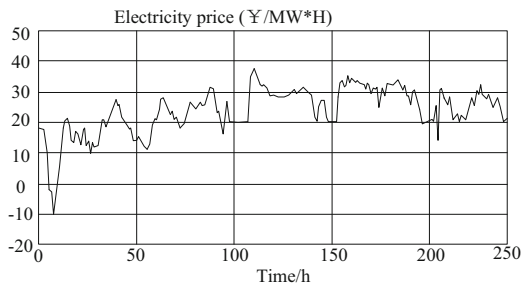
(a) Electricity price graph for one week in January



(b) Electricity price graph for one week in April



(c) Electricity price graph for one week in July



(d) Electricity price graph for one week in October

**Fig. 3.** Monthly clearing price of power plants in a certain area

times of the day. It can be seen that the electricity prices in this area fluctuate a lot in January and July and there are extremely high electricity prices.

Based on the monthly clearing price of the power plant shown in Fig. 3, three sets of forecasting methods are used to predict the monthly clearing price of electricity at different times in the above four months. The prediction results are as follows.

### 3.2 Experimental Result

#### The First Set of Experimental Results

Based on the experimental data selected in this experiment, three sets of forecasting methods are used to predict the monthly electricity price of the area in the first 10 days of January according to the monthly electricity price of the first 10 days of January as shown in Fig. 3 (a). The electricity price is cleared, and the average absolute percentage error in the formula (2) is used to calculate the relative error between the forecast results of the three sets of forecasting methods and the actual results of the monthly electricity clearing price in the region. The calculation results are shown in Table 3.

**Table 3.** Forecast results of the monthly clearing price of electricity in January (%)

Time/day	Method		
	Research method	Common method 1	Common method 2
1	0.01	2.23	0.84
2	0.05	4.35	0.30
3	0.07	4.90	0.87
4	0.09	2.34	0.71
5	0.06	2.27	0.49
6	0.03	2.25	0.49
7	0.02	2.66	0.91
8	0.04	2.59	0.55
9	0.03	2.49	0.59
10	0.02	2.44	0.27

As can be seen from Table 3, based on the monthly settlement electricity price in Fig. 3 (a) in early January, the method in this paper predicted the monthly settlement electricity price and the actual monthly settlement electricity price in late January in this region. The average absolute percentage error of the proposed method is less than 0.1%, which is obviously superior to the two traditional methods. This result shows that the method presented in this paper has a good prediction accuracy for monthly settlement electricity price.

### The Second Set of Experimental Results

Using three sets of forecasting methods, according to the monthly clearing price of electricity in the first 10 days of April shown in Fig. 3 (b), predicting the monthly clearing price of electricity for the 10 days after April in the area, and adopting formula (2) the average absolute percentage error, calculate the relative error between the prediction results of the three sets of forecasting methods and the actual results of the monthly clearing price of electricity in the region. The calculation results are shown in Table 4.

**Table 4.** Forecast results of monthly electricity clearing price in April (%)

Time/day	Method		
	Research method	Common method 1	Common method 2
1	4.31	8.88	9.15
2	4.43	8.15	9.35
3	4.54	8.10	9.13
4	4.69	8.59	9.27
5	4.75	8.64	9.91
6	4.72	8.52	9.22
7	4.05	8.59	9.84
8	4.67	8.55	9.73
9	4.51	8.61	9.69
10	4.48	8.60	9.76

It can be seen from Table 3 that the research method is based on the monthly clearing price of electricity in the first 10 days of April as shown in Fig. 3 (b), and predicting the monthly clearing price of electricity for the 10 days after April in this area, and the actual monthly clearing price. The average absolute percentage error of the cleared electricity price is less than 5%, which is significantly better than the two commonly used methods, and has better prediction accuracy of the monthly cleared electricity price.

### The Third Set of Experimental Results

Using three sets of forecasting methods, according to the monthly clearing price of electricity in the first 10 days of July shown in Fig. 3 (c), predicting the monthly clearing price of electricity for the 10 days after July in the area, and adopting formula (2) The average absolute percentage error  $P$  in, calculate the relative error between the prediction results of the three sets of forecasting methods and the actual results of the monthly clearing price of electricity in the area. The calculation results are shown in Table 5.

**Table 5.** Forecast results of electricity monthly clearing price in July (%)

Time/day	Method		
	Research method	Common method 1	Common method 2
1	1.49	4.98	6.98
2	1.45	4.94	6.94
3	2.81	6.30	8.30
4	2.35	5.84	7.84
5	2.72	6.21	8.21
6	2.59	6.08	8.08
7	2.79	6.28	8.28
8	2.03	5.52	7.52
9	1.55	5.04	7.04
10	1.37	4.86	6.86

It can be seen from Table 3 that the research method is based on the monthly clearing price of electricity in the first 10 days of July as shown in Fig. 3 (c), and predicting the monthly clearing price of electricity for the 10 days after July in this area, and the actual monthly clearing price. The average absolute percentage error of the cleared electricity price is less than 3%, which is significantly better than the two commonly used methods, and has better prediction accuracy of the monthly cleared electricity price.

#### The Fourth Set of Experimental Results

Using three sets of forecasting methods, according to the monthly clearing price of electricity in the first 10 days of October shown in Fig. 3 (d), predicting the monthly clearing price of electricity for the 10 days after October in the area, and adopting formula (2) The average absolute percentage error in, calculate the relative error between the forecast results of the three sets of forecasting methods and the actual results of the monthly clearing price of electricity in the region. The calculation results are shown in Table 6.

**Table 6.** October electricity monthly clearing price forecast results (%)

Time/day	Method		
	Research method	Common method 1	Common method 2
1	0.68	3.67	6.49
2	0.81	3.80	6.62
3	0.87	3.86	6.68
4	0.73	3.72	6.54
5	0.61	3.60	6.42
6	0.53	3.52	6.34
7	0.62	3.61	6.43
8	0.65	3.64	6.46
9	0.54	3.53	6.35
10	0.66	3.65	6.47

It can be seen from Table 3 that the research method is based on the monthly clearing price of electricity in the first 10 days of October as shown in Fig. 3 (d), and predicting the monthly clearing price of electricity for the next 10 days of October in this area, and the actual monthly clearing price. The average absolute percentage error of the clearing electricity price is less than 1%, which is significantly better than the two commonly used methods, and has better prediction accuracy of the monthly clearing electricity price.

## 4 Conclusion

This study fully analyzes the impact of the constant adjustment of supply and demand on the electricity price in the electricity market, and uses the time information and load information to construct the training sample of monthly settlement electricity price, thus combining with the BP neural network to complete the prediction of the monthly clearing price. The prediction accuracy of this method is high, and the average absolute percentage error of the prediction results is obviously less than that of the traditional method, which proves that it is suitable for popularization and application.

## References

1. Liao, Z., Chen, L., Huang, J., et al.: Medium and short-term electricity coal price forecast based on feature space transformation and LSTM. *J. Northeast. Univ. (Nat. Sci.)* **42**(4), 483–493 (2021)
2. Wei, Q., Chen, S., Huang, W., et al.: Forecasting method of clearing price in spot market by random forest regression. *Proc. CSEE* **41**(4), 1360–1367 (2021)
3. Liu, R., Jing, Z., Liu, Y.: Coupling clearing model of energy and reserve considering dynamic operation reserve demand curves. *Autom. Electr. Power Syst.* **45**(6), 34–42 (2021)
4. Cun, X., Qian, Z., Sun, Y., et al.: “Cost-Accuracy” hedging based load forecasting technique on two-stage electricity market. *Electr. Power* **53**(10), 172–179 (2020)
5. Gou, X., Xiao, X.: Short-term electricity price forecasting model based on empirical mode decomposition and LSTM neural network. *J. Xi’an Univ. Technol.* **36**(1), 129–134 (2020)
6. Zhao, S.: Research on regional hydropower market price forecasting model. *Water Conserv. Sci. Technol. Econ.* **26**(8), 5–9 (2020)
7. Huang, J., Huang, K., Xiang, D., et al.: Multi region electricity price prediction based on big data of electric power trading platform. *Electrotech. Appl.* **39**(9), 36–41 (2020)
8. Tian, D., Chen, Q.: Price prediction based on big data and Gaussian process. *Comput. Knowl. Technol.* **15**(19), 24–26 (2019)
9. Yao, N., Wang, X., Wu, T.: Prediction model of BP neural network based on online training. *J. Ili Normal Univ. (Nat. Sci. Ed.)* **13**(3), 64–69 (2019)
10. Zhao, S., Zhang, H.: LPG price forecast based on random forest. *E-science Technol. Appl.* **10**(3), 21–27 (2019)
11. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>

12. Liu, S., Liu, X., Wang, S., Muhammad, K.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
13. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>



# Automatic Response Method of Grid Demand Under the Background of New Energy Consumption

Wen-tao Xu<sup>(✉)</sup>, Chong-wei Ma, Guo-bin Wang, Ye-fei Li,  
and Hui Liu

Marketing Service Center (State Grid Ningxia Electric Power Co., Ltd. Metering Center), State Grid Ningxia Electric Power Co., Ltd., Ningxia 750000, China  
xuwentao@nx.sgcc.com.cn

**Abstract.** Smart grid demand response can reduce the peak power demand of the grid, improve the stability and reliability of grid operation, especially the ability of the grid to accept intermittent renewable energy generation through demand response. In view of the abnormalities in the analysis of user demand caused by the traditional automatic response method of power grid demand, which caused large fluctuations in electricity prices, the automatic response method of grid demand under the background of new energy consumption was designed. Constructing a power grid model as the research object, constructing a time-of-use electricity price model according to the characteristics of power grid operation, and setting demand response incentive methods. Analyze the characteristics of user power consumption and grid load, and design the automatic response algorithm for grid demand. At this point, the design of the automatic response method for grid demand in the context of new energy consumption has been completed. Construct an experimental link, and the experimental verification shows that the new response method can effectively control electricity price fluctuations and improve user satisfaction.

**Keywords:** New energy consumption · Grid demand response · Incentive mechanism · Smart grid

## 1 Introduction

At the beginning of the 21st century, our country has experienced a rapid development stage in which the whole society and economy have developed rapidly and the people's living standards have improved significantly. At the end of the "Twelfth Five-Year Plan", my country's economy has entered a shifting period, and economic growth has slowed down. The environmental pollution caused by the production and use of electric energy under the traditional extensive energy use method has been closely watched by the public. With the advent of a new era of economic development, the growth rate of electricity consumption in the whole society has slowed down, and the overall supply and demand of electricity exceeds demand. The rational development of electric energy and efficient and clean utilization will become one of the first tasks to achieve the goal of structural transformation under the new normal of the economy [1]. In order to

improve the power supply capacity of the power grid, it can be considered from two aspects: the transmission line of the power grid and the intelligent demand side response of the user side. Increase the current carrying capacity of the transmission line through the transmission line capacity increase technology, thereby improving the power supply capacity of the transmission line. By implementing demand response technology on the user side, researching smart load control technology on the residential side, effectively integrating user-side demand response resources, so as to improve the grid load curve, achieve peak shifting and valley filling, and improve the power supply capacity and operating efficiency of the grid.

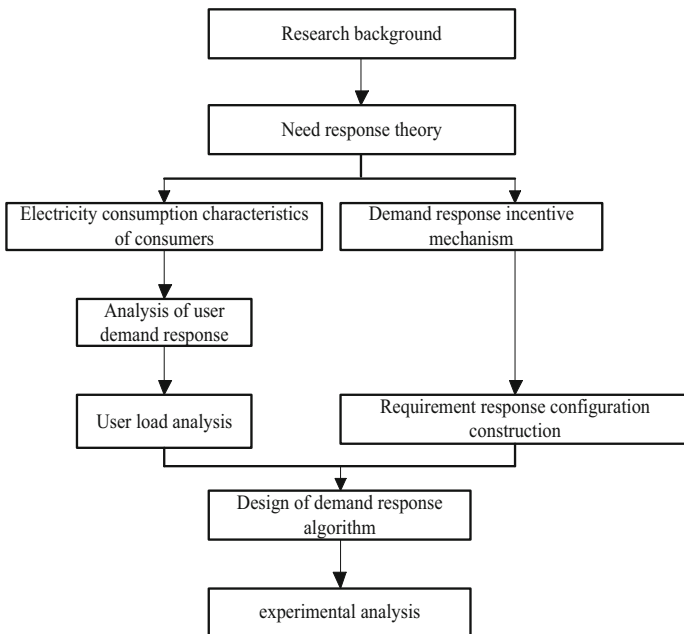
Demand response related theoretical ideas were introduced into my country in the late 1980s. After more than 30 years of experimentation and exploration, it has been integrated into the characteristics of my country's electricity market. Now the peak-valley time-of-use electricity price mechanism has been implemented in most parts of the country, and peak electricity prices Mechanisms and interruptible load mechanisms are still in the exploratory period [2]. Intelligent and orderly use of electricity belongs to the power demand side management technology, which refers to the load forecast of all electrical equipment participating in the demand response based on the previous load data when the load is tight. Control the power status of the equipment. However, the traditional demand-side management and control methods are generally semi-automatic or manual control. Measures such as power-limiting and opening the switch cannot achieve intelligent power consumption, and have a negative impact on the normal production and life of power users. The development of the smart grid provides opportunities for the development of advanced measurement technology, load intelligent control, and grid communication technology. While providing technical support for the smart grid, it also provides intelligent services for end power users. The construction of smart grid has brought opportunities for the development of demand response technology, and has produced a series of new load control methods and measures. The use of smart grid two-way interactive technology to achieve flexible control of demand response resources and load dispatch through two-way communication between users and the grid is the development trend of smart grids in the future [3].

Based on the above research background, this study carries out research on the automatic response method of power grid demand under the background of new energy consumption. In this study, on the basis of in-depth understanding of the existing research results at home and abroad, the basic theory of demand response and new energy development is analyzed, and the typical load curve and power consumption characteristics in the current life are analyzed. On this basis, the paper summarizes the impact of new energy consumption on user load and its implementation. Then consider the government departments, power grid companies, power companies, such as user participation main body the best cost performance, and use for reference the experience of other countries successfully cope with the demand, put forward in accordance with the current development environment of electricity market incentive mechanism, and through building incentive effect evaluation index system and evaluation model to complete the design of the demand for power grid automatic response method.



## 2 Design of Automatic Response Method to Grid Demand Under the Background of New Energy Consumption

Based on the study of automatic demand response methods, this paper proposes a new type of automatic demand response method for power grids based on smart technology and the characteristics of the power grid. This method combines Internet technology and automatic demand response to construct an automatic demand response method. Through the analysis of the most commonly used information transmission methods, the automatic demand response method design of the grid under the background of new energy consumption is carried out. The research idea of this paper is shown in Fig. 1.



**Fig. 1.** Technical research route

According to the above images, the structure of the full text is arranged as follows:

First, explain the background and significance of the research of the thesis, then introduce the current domestic and foreign research status of demand response, and finally give the main work and content arrangement of this research.

Then, the related theories used in this research are mainly introduced. Including the basic theory of demand response, as well as the characteristics and key technologies of user management. The theoretical knowledge in this link provides the basis for the subsequent research and design of automatic demand response methods.

Then, design demand response incentive mechanism and simulation model. This link first comprehensively analyzes the interest linkage relationship between the participating entities, and on this basis, proposes a reasonable incentive mechanism; then,

under the premise of comprehensively considering the costs and benefits of each participating entity, the incentive effect is constructed. Evaluation index system; Finally, the simulation analysis equation and model of the implementation effect of the incentive mechanism are proposed.

Finally, construct an experimental link to verify the effect of this method. Introduce the main research results of this research and the deficiencies in the research, and propose further research directions.

## 2.1 Grid Model Construction

There is no universally accepted unified definition of smart grid, and the focus of various definitions is different. According to the definition of the National Institute of Standards and Technology (KIST) [4], a smart grid is a network system that integrates a variety of digital computing, communication technologies and services into the power system infrastructure. It goes beyond the two-way flow of energy, two-way communication and control capabilities of smart meters in homes and businesses, and can bring new functions. The construction of smart grid can improve the robustness, self-healing ability and reliability of the grid. NIST proposed a conceptual model of the smart grid. This model supports the planning, requirements development, and documentation of the interconnected networks and equipment that make up the smart grid. NTST divides the smart grid into seven domains (including subdomains), including smart grid participants and applications. This conceptual model of the smart grid divides participants into participating equipment (such as smart meters and solar generators), systems (such as control systems), procedures, stakeholders who make decisions and exchange information to complete the application, as one or The application of tasks performed by multiple actors in a field (such as home automation, solar power generation, energy storage, and energy management). Based on the above content, the contents of the grid model construction in this study are set as follows (Table 1):

**Table 1.** Grid model elements

Element number	Element name	Content
1	Electricity users	Where electricity is used. Mainly for residential, commercial and industrial users. Participants can also generate, store and manage energy usage
2	Electricity market	The place where electricity is traded. Participants are operators and participants in the electricity market
3	Service provider	Provide service support for manufacturers, distributors and users. Participants are organizations that provide services to power users and utilities
4	Operator	Ensure the normal operation of the power system. Participants are managers of power transmission
5	Large power plant	Source of electricity. Participants are power plants that generate large amounts of electricity, and can also store energy for later distribution
6	Power transmission	It is a long-distance power operator that completes the batch transmission of electricity from generation to distribution, and can also store and generate electricity
7	Power distribution	Realize power transmission, consumption metering, distributed power generation and distributed storage to users. Participants are power distributors who manage the transmission of power to users or receive information from users

According to the content in the above table, combined with the “partial” modeling idea [5] to complete the grid model construction process. The “partial” modeling idea is to model various types of equipment and various energy systems separately. This type of modeling idea first performs detailed modeling of various equipment components, and then lists the corresponding energy balance equations according to the system energy type. The typical representative of this modeling idea is the smart grid modeling idea based on the energy bus architecture proposed in the literature. This idea first classifies the bus according to the energy transfer medium, and divides the equipment into the source and the conversion device according to the role of the energy conversion process equipment., Energy storage and load; then analyze and model various types of equipment respectively; finally, write the energy balance equations according to the energy balance of various types of buses. The focus of this modeling idea is the internal structure of the system and the conversion relationship of energy flow. Assuming that the power grid is composed of  $N$  nodes and  $l$  transmission lines, the DC power flow equation after dimensionality reduction (one node is selected as the reference node) can be expressed as:

$$BW = I \quad (1)$$

Among them,  $W$  is the  $(N - 1)$  dimensional node voltage phase angle vector, when  $W$  is known, the power flow on any transmission line can be calculated,  $I$  is the  $(N - 1)$  dimensional node injected power vector,  $B$  is the  $(N - 1)(N - 1)$  dimensional node voltage matrix, if  $\partial(n)$  Representing the set of neighbor nodes of node  $n$ , there are:

$$B_n = \begin{cases} \sum h_n & n \in s \\ -h_n & h \in \partial(n) \\ 0 & otherwise \end{cases} \quad (2)$$

where  $h_n$  is the voltage of the transmission line between node  $n - 1$  and node  $n$ . Use this model as the basis of this research.

## 2.2 Time-of-Use Electricity Price Model Construction

According to the new energy consumption requirements, the grid model constructed in the above article is the carrier, combined with the time-of-use electricity price mechanism [6], to construct a demand response model. The principle of the time-of-use electricity price mechanism is to divide it into three periods of peak, flat, and valley according to the changes in the daily load curve within a day, and set reasonable electricity price levels respectively. Compared with before the implementation of time-of-use electricity prices, user load will be adjusted according to the electricity price level of time-of-use electricity prices. The ratio of the load change over a period of time to the electricity price change is called the elasticity of demand. Obviously, the user’s response to the electricity price level includes not only the single-period response to the current period of electricity price and the multi-period response to other periods. Under the peak-valley time-of-use price mechanism, it is derived from the elasticity coefficient

and the cross-elasticity coefficient, which respectively represent the user's response to the electricity price. Single-period response and multi-period response.

Based on the principles of economics [7], the elasticity coefficient  $\alpha$  refers to the change in electricity demand caused by the fluctuation of the electricity price, and its mathematical value should be the ratio of the amount of electricity change to the amount of electricity price change, as shown below.

User's self-elasticity coefficient:

$$\alpha_{ii} = \frac{\Delta S_i / S_i}{\Delta D_i / D_i} \quad (3)$$

User's cross elasticity coefficient:

$$\alpha_{ij} = \frac{\Delta S_i / S_i}{\Delta D_j / D_j} \quad (4)$$

The calculation process of defining the amount of change in electricity consumption during a certain period of time is as follows:

$$\Delta S_i = \int [f_T(D_f, D_s, D_e) - f(D_i)] dt \quad (5)$$

$$\Delta D_i = D_T - D_i \quad (6)$$

In the formula:  $D_i$  represents the electricity price in the period  $i$  before the implementation of the peak-valley price;  $\Delta D_i$  represents the change in the electricity price of the user during the  $i$  period before and after the implementation of the peak and valley price;  $\Delta S_i$  represents the change in the electricity consumption of the user in the  $i$  period before and after the implementation of the peak-valley time-of-use price;  $D_T$  Represents the electricity price in the  $i$  period after the time-sharing management is implemented;  $D_f, D_s, D_e$  represents the electricity price in the peak, flat, and valley periods respectively;  $f_T(D_f, D_s, D_e)$  represents the load in the  $i$  period after the time-sharing management is implemented, which is a function of the electricity price in each period of peak, flat and valley;  $f(D_i)$  represents the implementation of the split When managing the user load in the first  $i$  time period, there is a functional relationship with the electricity price  $D_i$  at time  $i$ ; if  $\Delta D_i = 0$ , the self-elasticity coefficient of the user in the  $i$  time period is 0;  $\Delta D_j = 0$ , the cross-elasticity coefficient of the user in the  $i$  time period is 0. The electricity price elasticity matrix is expressed as:

$$E = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{bmatrix} \quad (7)$$

Electricity demand after user response:

$$\begin{bmatrix} D_1^* \\ D_2^* \\ \dots \\ D_n^* \end{bmatrix} = \frac{1}{n} \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \dots & \\ & & & D_n \end{bmatrix} E \begin{bmatrix} \Delta S_1/S_1 \\ \Delta S_2/S_2 \\ \dots \\ \Delta S_n/S_n \end{bmatrix} + \begin{bmatrix} D_1 \\ D_2 \\ \dots \\ D_n \end{bmatrix} \quad (8)$$

The above is a typical basic model of demand response based on the time-of-use electricity price mechanism. The main principle is that users adjust their own electricity consumption according to price changes, reduce electricity in the current period, and perform load transfer during cross-period periods. Through the demand elasticity matrix of formula (8), users Based on the load change and power after the electricity price change.

### 2.3 Demand Response Incentive Mode Setting

The user's willingness to participate in demand response projects is based on the premise that the benefit of demand response is greater than the cost of demand response. The benefit of demand response is closely related to the demand response incentive method. Specifically, it is mainly the type of demand response incentive, the price of power products, and the price of power products. There are three factors in government subsidies.

At present, demand response incentives are usually divided into two types: price-based demand response projects and incentive-based demand response projects [8]. These two types of projects have their own advantages and disadvantages: price-based demand response projects have most control signals. The control cycle is based on "hours". The power user's power load cannot be adjusted quickly according to the changes in the output of new energy, and it cannot ensure the economic and safe operation of the new energy after it is connected to the power system; the demand response project based on incentives is basically It can directly control the power load, can respond to the volatility and randomness of new energy in a timely and accurate manner, and respond to system signals, which has more advantages in the integration and consumption of new energy.

According to the analysis result of the incentive type, analyze the demand response method and complete the selection process. At present, there are two most widely used demand response methods: one is the time-of-use price, and the other is the interruptible load. The electricity price level during peak, flat, and valley periods has a key impact on users' electricity bills. Only when the saved electricity bills are greater than the loss caused by load reduction, can users take the initiative to adjust their electricity consumption behavior [9]. Based on the above analysis results, construct an incentive mechanism for maximizing benefits and evaluate its use effects. The specific evaluation indicators are as follows (Table 2):

**Table 2.** Evaluation indicators of demand response incentive mode

Evaluation index level	Evaluation index number	Evaluation index
I	1	Demand response participation
I	2	New energy access ratio
I	3	Avoidable peak load ratio
II	4	Load rate increase ratio
II	5	Benefit sharing ratio
III	6	Government subsidies
III	7	User Earnings Ratio
III	8	Grid Company Earnings to Cost Ratio
III	9	Profit-to-cost ratio of power generation companies
III	10	Emission reduction benefits

In order to objectively and reasonably evaluate the incentive effect of the user demand response incentive mechanism, the selected evaluation indicators must not only reflect the demand response benefits and demand response costs of various stakeholders, but also reflect the consumption of new energy. Factors such as the impact of the impact on the power system. Based on the above research and combined with the principle of indicator selection, the following indicators are proposed to evaluate and analyze the implementation effect of the incentive mechanism constructed in the previous section.

#### 2.4 Analysis of User Electricity Consumption Characteristics

According to the parameter requirements of the time-of-use electricity price model constructed above and the incentive mode, the user's electricity load is analyzed and used as the main content of the grid demand response. Electric load characteristic indexes are the basis for analyzing electric load characteristics. After long-term development, a complete load characteristic index system has been formed. These load characteristic indexes are commonly used in power grid planning and load forecasting. There are many indicators of power load characteristics, which are mainly divided into two types: numerical type and curve type. Some reflect the overall situation of power load characteristics, which are mainly used for horizontal comparison between domestic and foreign regions; some load characteristic indicators are mainly used for analysis and calculation of power system planning and design. In order to facilitate the analysis of load characteristics in practical applications, load characteristic indicators can be divided into three types: description, comparison and curve [10]. There are many power load characteristic indexes, and different load characteristic indexes have different functions, which reflect the different characteristics of electric load and are applied to different analysis needs. The following only analyzes and studies a few commonly used load characteristic indexes.

Daily load rate and daily minimum load rate:

Daily load rate and daily minimum load rate are used to describe the characteristics of the active daily load curve, indicating the unevenness of the power load distribution in a day, the daily load rate and daily minimum load rate and the nature, category, production shift and system of the power user. Factors such as the proportions of various types of electricity consumption in the country are related to, and at the same time, related to the measures to adjust the load. It is defined as follows:

Daily load rate and daily minimum load rate:

$$U_1 = \frac{E_{all}}{E_{max} \times 24} \times 100\% \quad (9)$$

$$U_{min} = \frac{E_{min}}{E_{max}} \times 100\% \quad (10)$$

In the above formula,  $E_{all}$  represents the total daily power consumption;  $E_{max}$  represents the maximum daily power consumption;  $E_{min}$  represents the minimum daily power consumption;  $U_1$  represents the daily load rate;  $U_{min}$  represents the minimum load rate.

Monthly load rate:

The monthly load rate is also called the monthly imbalance coefficient, which mainly reflects the influence of power users due to shutdowns, minor repairs of power equipment, or the addition of new power users. The change in the monthly load rate is mainly related to the composition of power users and the season of load. Sexual changes and holidays are closely related. The specific definition is as follows:

$$U_2 = \frac{M\bar{E}}{ME_{max}} \times 100\% \quad (11)$$

In the above formula,  $M\bar{E}$  represents the monthly average daily electricity consumption;  $ME_{max}$  represents the monthly maximum daily electricity consumption.

Seasonal load rate:

The quarterly load rate is also called the quarterly unbalanced coefficient, which is similar to the monthly load rate. It mainly reflects the seasonal changes in the electrical load of power users. For example, the factors that affect the quarterly load rate are mainly the seasonal configuration of electrical equipment and annual overhaul. Or the annual increase in power load, etc. The specific definition is as follows:

$$U_3 = \frac{ME_{max}}{Y_{max}} \quad (12)$$

In the above formula,  $ME_{max}$  represents the average value of the sum of the maximum monthly loads throughout the year;  $Y_{max}$  represents the maximum annual load.

Annual average daily load rate and annual load rate:

- (1) The annual average daily load rate mainly reflects the influence of the power load of the tertiary industry. The specific definition is as follows:

$$U_4 = \frac{\frac{1}{12} \sum ME_{\max}}{24 \times \frac{ME_{\max}}{12}} \times 100\% \quad (13)$$

In the above formula,  $U_4$  represents the annual average daily load rate.

- (2) The annual load rate is a comprehensive load characteristic index, which is mainly related to the changes in the power load structure of the tertiary industry. Under normal circumstances, the annual load rate will increase with the increase in the proportion of electricity consumption in the secondary industry. Rising, on the contrary, will decrease as the proportion of electricity consumption by the tertiary industry and household electricity consumption increases [11]. The specific definition is as follows:

$$U_5 = \frac{\bar{Y}}{Y_{\max}} \times 100\% \quad (14)$$

In the above formula,  $\bar{Y}$  represents the average annual load. Use the above formula to complete the calculation process of the load characteristics of the grid and grid users, and use this calculation result as the constraint condition of the grid demand response.

## 2.5 Demand Response Algorithm Setting

According to the analysis in the previous section, the essence of demand response for the optimization of social benefits is to make decisions for each power supply station/user within the scope of the power supply capacity of the power supply station and the user's basic power demand under the premise of meeting the basic constraints of power transmission. The power supply/power consumption is optimized to optimize social benefits. It can be seen that demand response should first ensure the safety and stability of the power grid, and then pursue the value of social benefits. If a power supply/power use decision-making scheme will cause one or more transmission lines to be overloaded and fail, then this scheme is unacceptable regardless of the social benefit value, because it will undermine the stability of the power grid [12]. On the contrary, if a power supply/power use decision-making scheme can avoid overloading of any transmission line in the transmission grid, then this scheme is acceptable even if the social benefit value generated by this scheme is not optimal [13]. In this study, the rapid demand response algorithm was used as the design basis to construct the demand response algorithm used in this study.



Set the initial search times  $a = 0$ , search factors  $\beta = 0$ ,  $\beta_i > \beta_j > 0$ , and search accuracy coefficient  $p > 0$ . Regardless of the power transmission constraints, calculate the output power  $\bar{k}_j$  of each power supply device  $j$  and the electric power  $\bar{g}_j$  of each user  $i$ , that is, solve the following convex optimization problem:

$$\begin{aligned} \min & \beta_2 \sum l_j(\bar{k}_j \eta) - \beta_1 \sum l_j(\bar{g}_j \eta) \\ & \sum \bar{k}_j = \sum \bar{g}_j \end{aligned} \quad (15)$$

Calculate the voltage phase angle  $\chi_n^t$  on each node  $n$  according to the following formula, and then calculate the power  $w_n^t$  flowing through each transmission line according to formula (15).

$$l_n \bar{k}_j - l_n \bar{g}_j - \sum (\chi_n^t - \chi_{n-1}^t) / u_{n,n-1} = 0 \quad (16)$$

According to the above formula, calculate the average violation degree  $Q$  of the transmission line load constraint:

$$Q = \frac{1}{N} \sum \max(|\chi_n^t| - \chi_n^{\max}, 0) \quad (17)$$

If  $Q = 0$ , and  $y = 0$  or  $\beta_u - \beta_l < e$ , Each power supply station  $j$  uses EGC to set its output power in this period to  $\bar{k}_j^t$ , and each user  $i$  uses ECC to set its power consumption to  $\bar{g}_i^t$ , and then exits the calculation. If  $Q > 0$ , then update  $\beta_l = \beta$  and calculate  $f_1(\beta_u, \beta_l, Q)$ ; otherwise, update  $\beta_u = \beta$  and calculate  $\beta = f_2(\beta_u, \beta_l)$ . Solve the following optimization problem, update the output power  $\bar{k}_j^t$  of each power supply station  $j$  and the electric power  $\bar{g}_i^t$  of each user  $i$ , then update  $y = y + 1$ , and turn to step 3. According to the content of the algorithm set above, the content set above is integrated into this calculation process, so as to realize the automatic response process of the grid demand. At this point, the design of the automatic response method for grid demand in the context of new energy consumption has been completed.

## 3 Experiment

### 3.1 Experimental Platform Construction

In this paper, an automatic response method to grid demand under the background of new energy consumption is proposed. In order to verify the effect of this method, in this study, an experimental link is constructed to analyze it. The simulation experiment form is used in this research to complete the experiment process. In order to achieve a high degree of simulation of the smart grid, CORE is used to simulate the operation of

the smart grid. CORE has scalability, customization, and a good human-machine interface. It can construct various types of virtual network scenarios, and run applications and network protocols submitted by users without modification to test its performance and feasibility. CORE can also establish connections with virtual networks constructed by other network simulation software or even real networks. The structure of CORE usage process is set as follows (Fig. 2):

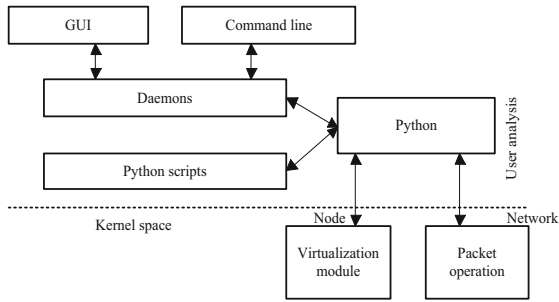


Fig. 2. CORE use process framework

Using the above framework to complete the simulation process of the smart grid, in order to test the feasibility of the method proposed in the article, a set of computer management software was developed for node configuration and visual analysis of the information collected by the coordinator device. Among them, the information collection subsystem The coordinator device transmits the collected information to the computer through the serial port, and uses this platform to complete the experiment comparison process.

### 3.2 Experimental Protocol Setting

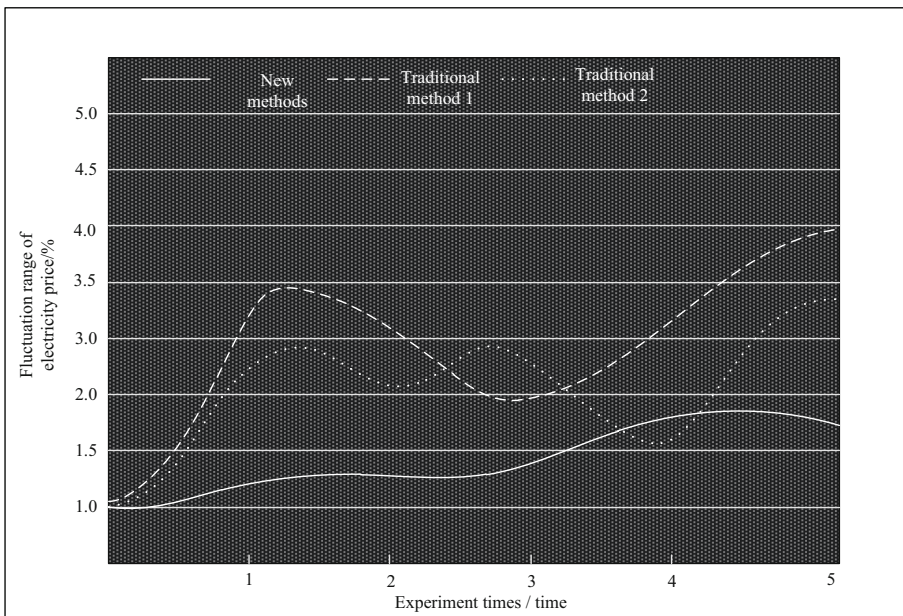
In the course of this experiment, the experiment comparison index is set as the economic benefit growth rate after application, the fluctuation range of power price and the stability of the grid load. The comparison process between the new automatic response method and the traditional method is completed by the experimental comparison index set above. In the experiment, CORE was used to build a smart grid model, the new automatic response method and the traditional method were used to complete the response process according to the grid demand, and the method comparison process was completed according to the indicators set above. In the experiment, each group of indicators was tested 5 times, and the use effect of the new response method and the performance difference of different response methods were determined based on the results of these 5 experiments.

### 3.3 Analysis of Results

**Table 3.** Economic benefit growth rate

Instructions	Comparison item	Calculation results/%
New response method	Highest economic benefit growth rate	5.14
	Minimum economic benefit growth rate	4.57
	Average economic benefit growth rate	4.89
Traditional method 1	Highest economic benefit growth rate	3.15
	Minimum economic benefit growth rate	2.27
	Average economic benefit growth rate	2.84
Traditional method 2	Highest economic benefit growth rate	2.15
	Minimum economic benefit growth rate	1.95
	Average economic benefit growth rate	2.07

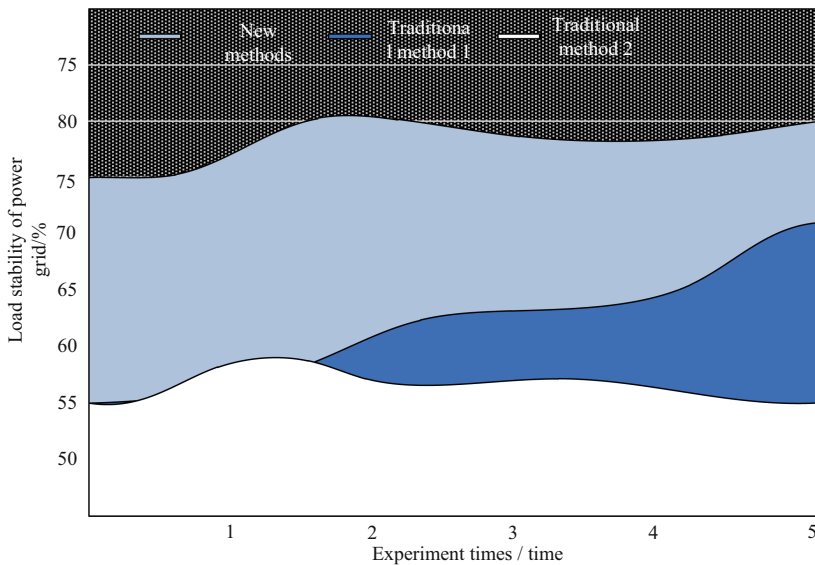
It can be seen from the above experimental results that both the new response method and the traditional response method can increase the economic benefits of power enterprises, but there are certain differences. The results show that this method can enable power companies to obtain higher economic growth rates and higher economic benefits, while ensuring the stability of the power grid and user satisfaction, while improving the economics of power grid operation. Compared with the new response method, the economic growth rate of the traditional method is relatively low. Under the premise of pursuing economic benefits, its use effect is not as good as the new response



**Fig. 3.** Electricity price fluctuation range

method. Based on the above analysis results, the new response method is more effective (Table 3 and Fig. 3).

After analyzing the above experimental results, the fluctuation of electricity prices after the application of different methods is determined. According to the analysis results, it can be determined that the new response method has a control effect on the stability of electricity prices, and after it is used, it can ensure that users accept electricity prices. Changes to avoid a decline in user satisfaction. The traditional method of electricity price fluctuates greatly. This situation will cause the problem of increased user dissatisfaction, which is unfavorable to the management and operation of power companies. Therefore, the new response method can be used as the main demand response method in the subsequent development of power enterprises (Fig. 4).



**Fig. 4.** Grid load stability

After analyzing the above experimental results, it can be seen that the new response method proposed in the article can effectively ensure the stability of power grid operation. After using this method, the operating state of the power grid can effectively ensure that users' power demand is met, while improving the power grid operation effect, reducing power price fluctuations, and improving users' evaluation of the power grid. Compared with the new response method, the traditional method cannot achieve the use effect of the new response method. The load rate of the power grid fluctuates greatly and the stability of the power grid operation is poor. Therefore, it can be determined that the new response method is more effective.

## 4 Discuss

Electricity demand response is that electricity users change their inherent electricity consumption patterns through electricity price signals or policy incentives in the electricity market, thereby adjusting load demand, optimizing electricity consumption behavior, saving energy, and optimizing resource allocation. At the same time, it contributes to the stability of the electricity market and the grid. Reliability has played an effective role in promoting. Based on the power demand side response, this research mainly studies three aspects: typical power user load characteristic classification technology, typical power user load characteristic analysis and power demand response algorithm. Based on the classification of power users' load characteristics and the analysis of typical power users' load characteristics, this research analyzes and studies the demand response of typical power users. In general, this topic has basically completed the scheduled tasks, but due to limited time and energy, there are still many areas that need to be improved. If you continue to conduct in-depth research, the main aspects are as follows:

The research in this paper uses numerical simulation to verify the effectiveness of the proposed method. To make the demand response method effective for the actual power grid system, it is necessary to build the actual physical environment of the simulated power grid and carry out physical simulation research. In this way, the proximity of the simulation environment and the actual working environment can be more reflected, and various factors of the power grid can be more fully and truly reflected. It is recommended to carry out physical simulation research in the follow-up research.

The demand response problem needs to consider the interests of both power supply companies and power users, which involves a number of indicators and parameters. When the demand response problem is transformed into an optimization problem, how to select the corresponding weight of each parameter remains to be further studied.

In future research, the energy optimization of typical power grids with distributed power sources and energy storage should be carried out to demand response models based on multi-dimensional grid systems to reflect continuity and integrity. Genetic algorithms should be used to simulate the bidding process, and only smart grids should be considered. The cooperation model for the non-cooperative model is not reflected. Future research should integrate the power grid control theory with demand response more deeply, explore the analysis process of the status of each user's main body on the demand side, and realize the optimization of the interests of each user main body.

## 5 Conclusion

In response to the energy crisis and environmental degradation, the construction and research of the Energy Internet is currently in full swing around the world. In recent years, my country has also issued a series of favorable policies to support and promote the construction of the Energy Internet. As the end energy supply system of the Energy Internet, the smart grid integrates a variety of energy sources and is equipped with various energy conversion and energy storage equipment. It is a complex

comprehensive energy utilization, conversion, storage and distribution system. In this study, the automatic response method of power grid demand under the background of new energy consumption was studied. In this study, the construction of power grid model as the research object, according to the characteristics of power grid operation to build a TOU price model, and set up the demand response incentive method. The characteristics of power consumption and power grid load are analyzed, and the automatic response algorithm of power grid demand is designed.

However, the method proposed in the article is implemented in a hypothetical future integrated energy market. The current energy market is independent, a multi-energy market has not yet been established, and the development of the future multi-energy market will be greatly affected by national policies, and it cannot be used in the actual energy market. The verification of the practicability and accuracy of the proposed model is still at the stage of theoretical analysis. Therefore, in the following research, in order to ensure the application effect of the method in this paper, the method in this paper should be put into use and further optimized through more data.

## References

1. Cao, W., Qi, B., Li, B., et al.: Balance and optimization of service in electric power communication network oriented to demand response. *Proc. CSEE* **40**(23), 7635–7643 (2020)
2. Wang, L., Wu, G., Zeng, S., et al.: The strategy of integrated demand response considering the interaction between distribution grid and industrial park. *Electr. Power Constr.* **40**(09), 52–63 (2019)
3. Hao, Y., Wang, P., Gao, C., et al.: Demand response mechanism of power grid based on energy efficiency of large user. *Electr. Power Autom. Equip.* **39**(04), 44–49 (2019)
4. Tang, Y., Zhao, S., Guo, Z., et al.: Demand response assisted photovoltaic absorption strategy considering the ramp rate limitation. *Renew. Energy Resour.* **38**(12), 1626–1632 (2020)
5. Sun, Y., Liu, D., Cui, X., et al.: Equal gradient iterative learning incentive strategy for accurate demand response of resident users. *Power Syst. Technol.* **43**(10), 3597–3605 (2019)
6. Lin, G., Lu, S., Guo, K., et al.: Stackelberg game based incentive pricing mechanism of demand response for power grid corporations. *Autom. Electr. Power Syst.* **44**(10), 59–67 (2020)
7. Liu, J., Wang, K., Ding, Z., et al.: Bus voltage-type demand response control for large-scale power shortage. *Proc. CSU-EPSA* **32**(03), 143–150 (2020)
8. Yao, G., Liu, X.: Hierarchical gaming for integrated energy system considered with renewable energies accommodation. *Comput. Eng. Appl.* **55**(07), 226–233 (2019)
9. Xue, J., Tang, Z., Sheng, R., et al.: New energy consumption model based on stackelberg game under the background of electricity market. *Mod. Electr. Power* **37**(03), 270–276 (2020)
10. He, Z., Zhou, G., Yang, Y.: Research on demand balance prediction of power supply system of photovoltaic power grid. *Comput. Simul.* **36**(03), 115–119 (2019)
11. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>

12. Liu, S., Liu, X., Wang, S., Muhammad, K.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
13. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>



# Effect Evaluation of College English Blended Teaching Based on Factor Analysis and Association Rules

Rong-rong Ruan<sup>1</sup> and Meng Ye<sup>2</sup>(✉)

<sup>1</sup> Fuyang Normal University, Fuyang 236037, China

<sup>2</sup> CSG Peak Shaving and Frequency Modulation Power Generation Co., Ltd.,  
Guangzhou 511400, China  
ym42856@163.com

**Abstract.** In order to effectively evaluate the effect of College English blended teaching and optimize the implementation effect of College English teaching. Based on factor analysis and association rules, this paper proposes a study on the effect evaluation of College English blended teaching. Starting from the practice of practical English blended teaching mode, according to the characteristics of practical English blended teaching mode, based on factor analysis and association rules, this paper evaluates the effect of College English blended teaching by combining practice with theory. Through investigation and analysis, it is proved that blended teaching mode can optimize college English classroom teaching and improve students' English practical ability. Further optimize the implementation effect of College English teaching.

**Keywords:** Factor analysis · Association rules · College English · Mixed teaching

## 1 Introduction

At the end of 1990s, foreign educational circles reflected on online learning and put forward the concept of blended learning, aiming at organically integrating traditional face-to-face teaching and online learning. Blended teaching is essentially a blended learning model, which is guided by cognitive learning, constructivism and other teaching theories, and uses a variety of teaching media, means and resources to closely combine classroom teaching with online teaching. In recent ten years, mixed teaching has attracted wide attention at home and abroad. Blended teaching is also a hot spot in foreign language education research. Many educators have integrated the concept of blended teaching into college English teaching practice, focusing on blended curriculum design and development, blended learning strategies, blended teaching practice mode and effectiveness research [2]. The research on the evaluation of mixed teaching effect mostly reconstructs the evaluation mechanism of teaching quality from the aspects of teachers, students and peers. The evaluation index system is constructed unilaterally from classroom or online teaching, which lacks the analysis of teaching practice, and only a few studies pay attention to the evaluation of mixed teaching of



individual courses, especially the evaluation of the effect of English audio-visual and speaking mixed teaching.

With the reform and innovation of the modern education system, practical English teaching mode is developing from single mode to mixed mode. Mixed teaching mainly focuses on students, guides students' independent and personalized development, and enables students to conduct independent practical learning in class and after class [3]. Practical English mainly cultivates students' ability of listening and speaking, and attaches great importance to the cultivation of students' knowledge and practical ability. After the implementation of the hybrid teaching model, practical English has been continuously innovating and developing. The research on the results of hybrid teaching is an important basis for promoting the development of teaching. After studying the practical implementation effect of mixed teaching, professional evaluation should also be conducted on the teaching results obtained to further promote the application of mixed teaching model [4]. The evaluation index system of blended teaching effect should fully reflect the advantages of online teaching and offline teaching. Evaluation of online teaching effect should include the number of students' visits, the number of discussion participants, the completion rate of task points, the situation of video viewing, and the average score of homework, etc. [5]. Offline teaching effect evaluation includes three indicators: academic performance comparison, student evaluation of teaching, and peer evaluation. It focuses on the evaluation of teachers' teaching attitude, teaching skills, teaching methods, classroom organization, depth and breadth of teaching content, and students' satisfaction with teachers. In order to effectively evaluate the effect of College English blended teaching and guide students' self-development and personalized development, this paper proposes a study on the evaluation of College English blended teaching effect based on factor analysis and association rules. This paper constructs the structure system of College English blended teaching, emphasizes the organic combination of online and offline and the main role of students, evaluates students from a dynamic perspective and way of thinking, and obtains comprehensive indicators (principal components) and original variables by using factor analysis statistical method.

## **2 Evaluation of Blended Teaching in College English**

### **2.1 The Structure System of College English Blended Teaching**

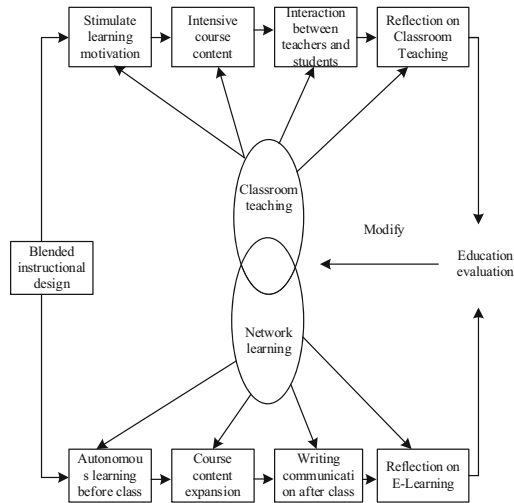
Hybrid teaching is a combination of online and face-to-face. Compared with simple face-to-face or online learning, it not only reflects the different learning methods, but also reflects the different ideas. Blended teaching reflects the spiral development of learning concept [6]. We should not unilaterally emphasize the application of online learning environment and ignore the traditional face-to-face teaching, nor emphasize student-centered and ignore the leading role of teachers, or emphasize one kind of learning theory and ignore the guiding role of other learning theories. For the investigation of classroom teaching has a relatively mature framework [7]. According to the teaching process of management by objectives, teaching is divided into three stages: teaching preparation, teaching implementation and teaching evaluation. Teaching

preparation refers to the formulation of teaching plan before classroom teaching. Teaching implementation is divided into three types: main teaching behavior, auxiliary teaching behavior and classroom management behaviorism. Teaching evaluation refers to the evaluation of students' academic achievement and teachers' classroom teaching after classroom teaching. Based on this, the paper standardizes the dimensions of College English teaching implementation behavior (Table 1):

**Table 1.** The dimensions of College English teaching implementation behavior

The dimension of teaching implementation behavior				
Teaching behavior				
Main teaching behaviors				
Presentation behavior	Dialogic behavior	Guiding behavior	Auxiliary teaching behavior	Classroom management behavior
Text presentation	Discuss	Cooperative learning guidance	The application of classroom reinforcement technology	The management of classroom problem behavior
Language presentation	Q & A	Guidance of autonomous learning	Stimulation and cultivation of learning motivation	Classroom problem behavior
Action presentation	–	–	The construction of good classroom atmosphere	–
Audio visual presentation	–	Inquiry learning guidance	The realization of teachers' expectations effect	Prevention of classroom problem behavior

Under the mixed teaching environment, teachers' teaching is not limited to the classroom. Some researchers have explored the framework of teaching process under the blended teaching mode. The process of blended teaching is divided into four links: course introduction, activity organization, learning support and teaching evaluation. Curriculum introduction refers to the exchange of curriculum requirements and arrangements between teachers and students, and the guidance of learning strategies. Activity organization refers to three forms of activities: class collective learning, group cooperative learning and individual learning. Learning support means that students get help to solve learning difficulties, which can come from teachers or classmates [9]. Teaching evaluation refers to checking and evaluating the expected teaching effect. The characteristics of blended teaching make its evaluation methods richer. Different learning forms have different learning processes and produce different learning results, so it is necessary to use a variety of evaluation methods. Based on this, the framework of mixed teaching process is further constructed (Fig. 1).



**Fig. 1.** Optimization of College English blended teaching process framework

Like traditional teaching, the framework divides the process of blended teaching into teaching preparation, teaching implementation and teaching evaluation. Teaching preparation in the mixed teaching environment, that is, teaching design, should consider both classroom face-to-face teaching and online learning. Teaching implementation is divided into face-to-face classroom teaching and online learning. In the classroom teaching part, teachers' main behaviors include stimulating students' learning motivation; Elaborate the course content and highlight the key and difficult points; Teachers and students interact and communicate to promote the construction of meaning; Teachers reflect on classroom teaching. In online learning, teachers' main behaviors include providing resources. Organize online interactive communication. Reflection on online learning. Teaching evaluation includes the evaluation of face-to-face classroom teaching and online learning. This framework follows the classic analytical framework of teaching preparation, implementation and evaluation, which fully reflects the characteristics of blended teaching, but the characteristics of personalized teaching emphasized by blended teaching are not clearly reflected in this framework [10]. The implementation is divided into the following four dimensions: building a mixed learning environment; Building mixed learning resources; Design and organize mixed learning activities; Implementing blended learning evaluation. This paper explores the structure of college English blended teaching. The following framework highlights the emphasis on the organic integration of online and offline and the main role of students. The framework is constructed from the perspective of students, but there is no clear explanation for teachers' behavior and role (Fig. 2).

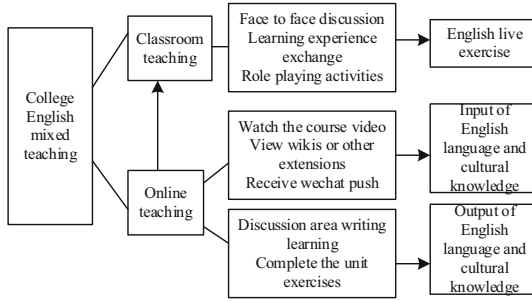


Fig. 2. Influencing factors of College English blended teaching

The reform of classroom teaching methods by teachers will be affected by internal factors, external factors and situation factors [11]. Internal factors refer to teachers’ own factors, such as growth experience, learning experience, teaching experience, etc. External factors refer to macro-social and cultural environment, such as economic development, foreign language education policies, etc. Situational factors mainly refer to the school environment in which teachers live, including school culture, teaching management, teaching resources, student sources, information technology, etc. Hybrid university English teaching factors affecting structure system not only be used to explain the factors affecting foreign language teachers classroom teaching method reform, can also be applied to analyze other situation of teacher's teaching behavior influence factor, the study found the teacher's choice of teaching strategies and use will be affected by factors of the students, curriculum factors and environmental factors, as well as the concept of teachers and the teacher individual characteristic factors. The student factor is also an important aspect that affects the teacher's teaching. If students have a good foundation and participate actively, teachers tend to implement student-centered teaching; On the contrary, if students are more passive and not actively participate in classroom activities, teachers will adopt teacher-centered teaching mode. In addition, good teacher-student relationship is also an important factor to promote teachers’ teaching. Let the linear regression model of random variable  $y$  and general variable  $x$  be as follows:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + \varepsilon \tag{1}$$

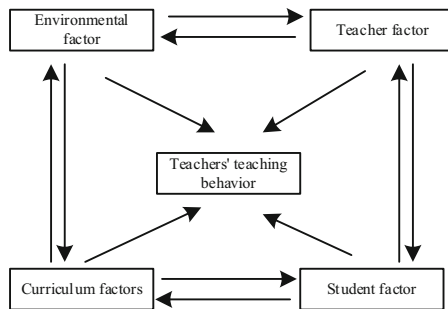
Based on this, the linear regression model can be expressed as:

$$\begin{cases} y_1 = \beta_0 + \beta_1x_{11} + \beta_2x_{12} + \dots + \beta_px_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1x_{21} + \beta_2x_{22} + \dots + \beta_px_{2p} + \varepsilon_2 \\ \dots \\ y_n = \beta_0 + \beta_1x_{n1} + \beta_2x_{n2} + \dots + \beta_px_{np} + \varepsilon_n \end{cases} \tag{2}$$

In matrix form:

$$Y = X\beta + y\varepsilon \tag{3}$$

To sum up, teachers’ teaching behavior is one of the decisive factors affecting teaching results, and the factors affecting teaching behavior can be summarized as follows: teachers’ own factors. Teachers’ teaching behavior is influenced by their own teaching ideas, teaching ability and teaching tact, which are influenced by teachers’ growth experience, educational background, learning experience and other factors. The nature of courses, class size, class hours arrangement, teaching objectives, teaching contents, teaching modes, class time and so on will have an impact on teaching behavior. Environmental factors can be divided into external environmental factors and internal environmental factors [12]. Teachers’ teaching is based on schools, so environmental factors outside schools are classified as external environment, such as social, political and economic conditions, educational policies, educational reform, teachers’ family factors and so on. All aspects of the school environment are classified as internal environment, including the school’s material conditions, policies, atmosphere and teaching management. These four factors do not work in isolation, but interact with each other, and have a comprehensive impact on teachers’ teaching. The following figure more intuitively shows the interaction between various factors and the influencing factors on English teaching evaluation (Fig. 3).



**Fig. 3.** Influencing factors of English teaching evaluation

To sum up, in terms of physical space, hybrid teaching includes online platform and offline classroom. On the online platform, students can learn autonomously and interact with teachers and students; In the face-to-face class, students show their achievements, teachers give feedback and organize discussions. Teaching evaluation runs through the two models.

### 2.2 Evaluation Index of Blended Teaching in College English

The evaluation of blended teaching should not only be based on one or two summative indicators, but also through a dynamic perspective and way of thinking to evaluate the students [13]. In the dynamic evaluation index, it can timely reflect the students’

learning status, such as specific to a week, a day. In this way, teachers can track and supervise the learning process of students according to the dynamic indicators, remind learners to pay attention to adjust the learning skills, and solve the problems encountered by students in time [14]. The comprehensiveness of the evaluation content requires that in the evaluation index, the investigation of knowledge points should be comprehensive, especially in the mid-term and final examination, and the distribution of knowledge should be comprehensive, not just around a certain part. At the same time, for the situation of students' participation in the discussion, we should not only examine whether the students participate in the discussion, but also examine what role they play in the discussion and their contribution to the team. The results of evaluation must be beneficial and set up to promote students' learning effect [15]. Evaluation is not only for students to summarize, but also for students to find their own shortcomings through evaluation, so as to improve the learning effect purposefully and directionally in the next study. Based on the above principles and the availability of data, the relevant indicators of the evaluation system and the corresponding evaluation methods, data types and evaluation subjects are as follows (Table 2):

**Table 2.** Evaluation system and index description table

Evaluation type	First level indicators	Secondary indicators	Evaluation method	Evaluation subject	Data type
Summative evaluation	Final exam	Final exam	Difference test	Teacher	Objective data
	Students' subjectivity	Satisfaction	Descriptive statistics	Student	Subjective factors
		Learning style	Descriptive statistics	Student	Subjective factors
Process evaluation	Pre class learning evaluation	Number of preview videos per person per month	Difference test	Student	Objective data
		Preview video viewing time per person per month	Difference test	Student	Objective data
	Evaluation of classroom activities	Number of discussions per person per month	Difference test	Student	Objective data
		Number of questions per person per month	Difference test	Student	Objective data
	After class learning evaluation	5 assignments and mid-term examination results	Difference test	Teacher	Objective data
The combination of process evaluation and summative evaluation	Construction of evaluation model	Combining process evaluation index with summative evaluation index	Correlation analysis	Combination of subjective data and objective data	
			Principal component analysis	Combination of subjective data and objective data	
			Regression analysis	Combination of subjective data and objective data	

The process is adopted in this paper can be seen from the table above six evaluation indexes, four summative evaluation index to evaluate, emphasizes the process assessment and summative assessment, evaluation indexes and involve wide, there are both subjective data and objective data, there are both students and teachers in the evaluation subject, evaluation system composed of these factors obviously satisfy the above principles, namely the evaluation criterion of differentiation, The diversification of the evaluation subject, the diversification of the evaluation method and the comprehensive evaluation subject. For dynamic evaluation of mixed teaching, the data collected in online learning can be used. However, considering the actual situation of operation in the random process in this paper, the course has fewer hours, so the dynamic data can be replaced by node data. The evaluation method in this paper also meets the dynamic evaluation process of the fifth point. It is assumed that the evaluation population is normally distributed.

If  $(x_{11}, x_{21}), (x_{12}, x_{22}), \dots, (x_{1n}, x_{2n})$  is a pair of samples with a volume of  $n$  selected from two populations, where  $x_{11}$  is the first pair of values in the sample, one of the two population samples comes from population 1 and the other comes from population 2, the difference value of each pair of sample data is calculated, and results are obtained. The difference sample is expressed as follows:

$$\begin{aligned} x_{11} - x_{21} &= d_1, x_{12} - x_{22} \\ &= d_2, \dots, x \end{aligned} \tag{4}$$

Its mean and standard deviation are respectively:

$$\bar{d} = \frac{\sum_{j=1}^n d_j}{n(x_{11} - x_{21})} \tag{5}$$

$$s_d = \sqrt{\frac{\sum_{j=1}^n (d_j - \bar{d})^2}{n - 1}} \tag{6}$$

Similarly, the mean and standard deviation of the population are:

$$\mu_d = \frac{\sum_{j=1}^N d_j}{N} = \mu_1 - \mu_2 \tag{7}$$

$$\sigma_d = \sqrt{\frac{\sum_{j=1}^N (d_j - \mu_d)^2}{N}} \tag{8}$$

The process evaluation is mainly divided into three evaluation nodes, the evaluation of pre-class learning, the evaluation of classroom activities and the evaluation of after-class

learning [16]. Classroom learning evaluation includes the monthly viewing times of preview videos per person, the monthly viewing time of preview videos per person, and the answer scores of online preview. The evaluation of classroom activities mainly includes the number of discussions and questions each student takes every month [17]. The evaluation indexes of classroom activities are the scores of monthly exams and midterm exams, and the number of self-study sessions each student takes every month. The changes of these indicators in four months were analyzed and compared, and the number of videos per person watched per month was taken as the number of preview times of students in mixed class and the number of preview times of students in traditional mixed class to conduct a difference test. If the population of differences is normally distributed, then the sampling distribution is also normally distributed [18]. The mean and standard deviation of the sampling distribution of mean difference are:

$$\mu_{\bar{d}} = \sigma_d \mu_d = \sigma_d(\mu_1 - \mu_2) \tag{9}$$

$$\sigma_{\bar{d}} = \frac{\sigma_d}{\sqrt{n}} \tag{10}$$

Since the mean and standard deviation of the population are unknown, distribution or Z distribution can be used for analysis [19]. When two populations do not obey normal distribution, Wilcoxon (W) rank sum test is used. There are two independent samples X and y, the sample size is n1 and n2 respectively, and the rank sum is wx and wy respectively:

$$W_x + W_y = 1 = 2 + \dots + n - \mu_{\bar{d}} = \frac{n(n+1)}{2} - \mu_{\bar{d}} \tag{11}$$

If:

$$W_1 = W_x - \frac{n_1(n_1+1)}{2} \tag{12}$$

$$W_2 = W_y - \frac{n_2(n_2+1)}{2} \tag{13}$$

And using standard normal distribution Z to approximate test, we can get:

$$\begin{aligned} Z &= \frac{W_x - \mu \pm 0.5}{\sigma(W_x + W_y)(W_2 - W_1)} \\ &= \frac{W_x - \frac{n_1(n_1+n_2+1)}{2(W_x+W_y)(W_2-W_1)} \pm 0.5}{\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12} - \frac{n_1n_2 \sum (\tau_j^3 - \tau)}{12(n_1+n_2)(n_1+n_2-1)}}} \end{aligned} \tag{14}$$



The mathematical model of factor analysis is as follows:

$$X = AF + \varepsilon \tag{15}$$

where  $X$  is the original variable and  $F$  is the common factor variable,  $\varepsilon$  is a special factor, which means the part of the original variable that cannot be explained by the common factor variable. It is equivalent to the residual part in multiple regression analysis.  $A$  is the factor load matrix, and  $a_{pm}$  is the correlation coefficient between the  $p$  th original variable and the  $m$  th common factor variable. The specific expression is as follows:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{pmatrix} \tag{16}$$

Based on the above algorithm, the comprehensive indicators (principal components) obtained by using the statistical method of factor analysis have the following characteristics with the original variables: each principal component can represent a linear combination composed of the original indicators; secondly [20], there should be no correlation between principal components; then the number of principal components should obviously be less than the number of original variables; finally, the principal components should cover most of the information of the original variables [21]. Through principal component analysis, the evaluation index can be reduced in dimension, and the main index can be extracted. According to factor analysis, the influence coefficient of each principal component and final grade can be obtained. The principal component with the highest coefficient is the most important factor affecting the learning effect (Table 3).

**Table 3.** Detailed description table of evaluation index

Index	Index description
Participation in the discussion	The full score is 10
Online answer results	A total of 10 multiple-choice Questions, the difficulty of these multiple-choice questions is general, 1 point for each question, the full score is 10 points
Preview video viewing duration	Preview video viewing duration = online duration - Online answer duration
Preview times before class	Whether or not to log in to the online learning platform every week, and online for up to 10 min is recorded as a preview for the student in that week
Homework results	That is, the score of the paper surface, each assignment is similar to the form of the test paper, with a full score of 100
Number of classroom questions	Every time you ask a question in class, it will be 1 point, and no question will be 0 points
Mid term and final examination results	And paper results

Based on the above evaluation system, college English teachers can not only control their own teaching process, understand the effects of all aspects of teaching, and adjust teaching design in time, but also fully understand students' learning status, urge and motivate students' learning participation, and keep students' interest and enthusiasm in learning for a long time [22]. Today, with the rapid development of information technology, college English teachers should make full use of Internet information technology, design evaluation dimensions and methods that are suitable for all aspects of teaching activities, and make a comprehensive and objective evaluation of learners' learning performance at all stages, so as to ensure that college English blended teaching is more efficient and quality.

### 2.3 The Realization of College English Mixed Teaching Evaluation

Teaching evaluation is the process of studying the value of teachers' teaching and students' learning, which plays a positive role in improving the teaching effect. In order to effectively improve the teaching effect and promote the positive role of network classroom in higher education, a set of scientific assessment and evaluation system should be constructed in the mixed teaching mode. Based on the theoretical and practical investigation in the early stage, a process-based assessment and evaluation system is constructed [23]. The specific framework is shown in Fig. 4. The assessment methods in this assessment and evaluation system can be selected and adjusted according to different teaching contents.

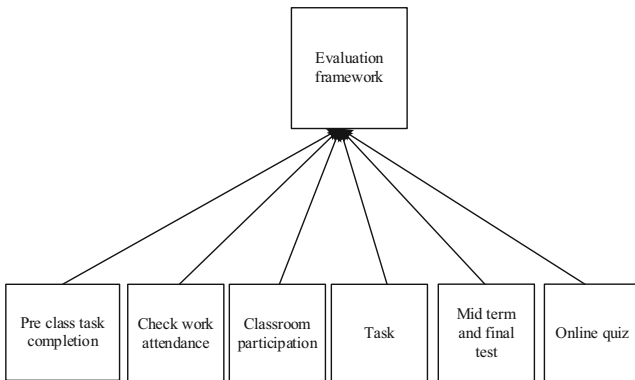


Fig. 4. A framework for grade evaluation of blended teaching model

Hybrid teaching can be divided into network course learning and face-to-face interpretation and practice of two phases, aims to guide students to acquire knowledge actively, and guide its applying theoretical knowledge to practice, therefore, the hybrid teaching evaluation content mainly examine the degree of knowledge acquisition, students ability of the application level and the degree of the practice of theoretical knowledge. Different from the traditional evaluation system, which focuses on the

basic theory and knowledge, the evaluation content of blended teaching is more comprehensive, including the evaluation of online course content, the specific performance of group cooperation, the attitude towards task activities and the comprehensive ability performance of offline practice activities. Among them, individual performance, such as class participation attitude, positive or not, and the completion of the comprehensive task in the network classroom, are the subjective motivation for the assessment of learning subjects. The completion of group discussion and collaboration, achievement report and display, and comments outside the group is used to examine students' practical application ability of knowledge, cooperation and communication ability, and language organization and expression ability. The blended teaching based on the network course will realize the transformation of the evaluation system from "theoretical cognition" to "ability application".

Teaching evaluation is the most important link in teaching design, which emphasizes taking teaching target as the main reference object. The goal of college English teaching in China is to combine basic grammar knowledge, listening, speaking, reading, writing and translation skills with language culture organically, to cultivate students' comprehensive ability to use language, and to cultivate students' independent learning ability, exploration ability, international vision and cultural accomplishment in practical teaching. The overall goal of teaching activities includes cognitive field, emotional field and psychological field. In order to achieve the final goal of each field, it is necessary to set a series of goals from low to high. This model is based on Bloom's theory of categorical teaching objectives, which divides blended teaching objectives into three aspects: cognitive objectives, skill objectives and affective objectives. These three objectives constitute the core literacy of college English learners and run through the whole semester. In blended teaching, when evaluating college English courses, these core qualities should be quantified one by one, and specific evaluation items and methods should be adopted to form a complete teaching design and reflect the diversity of the evaluation system. Based on this, the evaluation system of blended teaching should be optimized as follows (Table 4):

**Table 4.** The Evaluation System of College English Mixed Teaching

Core literacy	Evaluation project	Concrete content	Evaluation method
Skill goal (30 points)	Learning files (10 points)	Self learning records of MOOC resources and Learning Websites	Students' self-evaluation, mutual evaluation and teachers' evaluation
	Term project (10 points)	Individual or group to complete English short play performance, etc	Teacher evaluation
	Use of mobile tools (10 points)	Use wechat, microblog and other mobile tools to receive and share information	Students' self-evaluation, mutual evaluation and teachers' evaluation

(continued)

**Table 4.** (continued)

Core literacy	Evaluation project	Concrete content	Evaluation method
Cognitive goal (40 points)	Classroom test and final written examination (30 points)	Vocabulary and grammar, writing, listening, reading comprehension, translation	Teacher evaluation
	Textbook learning (10 points)	Preview before class	Teacher evaluation
Emotional goals	Cultural literacy and international vision (10 points)	Understand Chinese traditional culture and be able to communicate in English	Students' self-evaluation, mutual evaluation and teachers' evaluation
	Effective learning strategies and behaviors (10 points)	Perseverance, willingness, patience, concentration, etc	Students' self-evaluation, mutual evaluation and teachers' evaluation
	Cooperation and innovation ability (10 points)	Group cooperative learning ability, innovative thinking ability	Students' self-evaluation, mutual evaluation and teachers' evaluation

By combining the process evaluation data and the summative evaluation data, the correlation analysis is conducted based on the final exam results. If the indicators in the process evaluation or the summative evaluation are significantly correlated with the results, it can be shown that the selection of research indicators and the construction of research system in this paper are better.

### 3 Analysis of Experimental Results

In order to verify the evaluation effect of College English Hybrid Teaching Based on factor analysis and association rules, the paper makes an empirical analysis, and takes the super star network teaching platform and learning app as the supporting environment for hybrid teaching. The system is composed of chapters, materials, assignments, evaluation, testing, notification modules, and provides a space for network communication between teachers and students. Actively strengthen the interaction and learning guidance between teachers and students online. The teacher asks students to check the teaching contents of relevant chapters and complete the micro course task points in relevant chapters. Teachers will arrange network homework regularly, check online autonomous learning in time, and include the results into the semester results in a certain proportion. Among them, group A is the evaluation method of this paper, and group B is the traditional evaluation method.

**Table 5.** Online learning situation table

Group	Visits	Number of discussions	Number of people	Task point completion rate	Total video viewing time	Average video viewing duration	Homework average
A	4883	432	30	98%	2873.9 min	93.75 min	97.68
B	1648	139	29	58.62%	979.8 min	33.78 min	83.13

It can be seen from Table 5 that Group A is obviously superior to Group B in terms of the number of visits, the number of participants in discussions, the completion rate of task points, the video viewing situation, the average score of homework, etc., and the learning effect between Group A and Group B is significantly different. It can be seen that in the network environment, in the class with mixed reform, students are more motivated to learn independently after class, have better participation in discussion, have higher completion rate of task points and videos, and have better completion quality of homework. This is mainly due to the fact that under the guidance of teachers, group A students develop the habit of self-study on the Internet, can actively consult relevant materials, make full use of online listening training, strengthen language input, and enhance the interaction between teachers and students, while group B students have insufficient motivation for self-study on the Internet because their teachers have not made rigid requirements, and log in to the platform only to complete their homework, while watching micro-lesson videos and network interaction are not ideal. In the classroom practice, the pre-test and post-test were carried out for the students in group A and B, and the SPSS software was used for comparative analysis. In the classroom, we carefully observe students' reactions, interview students' feelings and record students' changes, aiming at verifying the theory with practice and drawing effective conclusions (Table 6):

**Table 6.** Pre-test English teaching scores

Achievement	Group	Number of people	Mean value	Standard deviation	Identification error of mean value
	A	30	66.79	13.35	2.32
	B	29	66.33	13.85	2.54

According to the table, the average scores of English listening, listening and speaking in both groups are 65.78 and 65.32, respectively, among which the scores of Group A are slightly better than those of Group B, and the average difference between the two groups is not significant, and the English listening level of both groups is basically at the same level.

**Table 7.** Post-test results

Achievement	Group	Number of people	Mean value	Standard deviation	Identification error of mean value
	A	30	86.05	5.92	1.06
	B	29	79.64	6.51	0.15

It can be seen from Table 7 that after the implementation of the mixed teaching mode, the average post-test scores of Group A and Group B are 86.01 and 79.74, respectively, and the scores of Group A are better than those of Group B, with obvious difference in average. Thus, after receiving mixed teaching, the results of Group A are higher than those of Group B, which also proves that mixed teaching plays a positive role in English audio-visual teaching and has a positive impact on students' listening and speaking ability. These data show that most students agree that mixed teaching can greatly promote their English learning; During the period of mixed teaching, when students use the network to study independently, teachers' guidance and help are very necessary, which also confirms the importance of adopting mixed teaching. The course group set up an open demonstration class of English audio-visual, speaking and held a teaching salon of mixed curriculum reform to share the experience of mixed reform with peers, which won unanimous praise. The course group held a teacher symposium, and invited teachers to evaluate the course from five aspects: teacher quality, teaching content, teaching condition, teaching method and teaching effect (Table 8).

**Table 8.** Peer evaluation form

Index	Teaching content (20)	Teaching conditions (20)	Quality of Teachers (20)	Teaching effect (20)	Teaching methods (20)	Total 100
Score	18	19	19	18	18	92

Teachers participating in the evaluation generally believe that the course teaching has distinctive features, can make full use of multimedia means, and has various content forms and flexible teaching methods. By adopting the flip classroom mode, students can watch micro-lessons in advance and complete the task points. Students can absorb knowledge according to their own characteristics and rhythm, and then return to class discussion after completing the comprehensive study of knowledge, which can significantly improve the quality of discussion and greatly improve the degree of interaction between teachers and students. Extracurricular online self-study enhances students' learning enthusiasm and initiative, effectively improves learners' listening and speaking ability and cultivates their English skills.

## 4 Conclusion

Hybrid teaching model applied to the English language teaching, pay more attention to training students' language application ability, cooperation, explore, autonomous learning and so on the many kinds of teaching methods, teachers as the guide and students as the main body, let English teaching activities to achieve learn and transformation of knowledge, make the teaching process to realize from the focus on the teaching goal to the student individuality development needs. Mixed teaching requires the participation and cooperation of teachers and students, which is based on practice teaching to jointly participate in the exploration and research of mixed teaching mode. At present, practical blended English teaching in colleges and universities is still in the stage of active exploration and needs to be persisted for a long time. In the future development, we should further optimize the process framework of College English blended teaching and add summative evaluation indicators to evaluate the effect of College English blended teaching.

**Funding.** In 2018, the key project of humanities and social sciences research in Anhui colleges and universities "Research on the construction of debate style based on dialogue syntax theory" phased results (project number: SK2018A0277); The 2018 Fuyang Normal University Humanities and Social Science Research General Project "The Cognitive Pragmatic Study of English Debate" (project number: 2018FSSK06).

## References

1. Dou, Q.: Multimodal discourse analysis in the blended teaching of college English flipped class. *Int. J. Electr. Eng. Educ.* **11**(5), 002072092110042 (2021)
2. Yang, C., Huan, S., Yang, Y.: Application of big data technology in blended teaching of college students: a case study on rain classroom. *Int. J. Emerg. Technol. Learn. (iJET)* **15**(11), 4 (2020)
3. Wang, R.: Massive open online course platform blended english teaching method based on model-view-controller framework. *Int. J. Emerg. Technol. Learn. (iJET)* **14**(16), 188 (2019)
4. Bai, X., Gu, X.: Group differences of teaching presence, social presence, and cognitive presence in a xMOOC-based blended course. *Int. J. Distance Educ. Technol.* **19**(2), 1–14 (2021)
5. Malinee, V.V., Senthamarai, T.: The use of Web 2.0 tools in English for specific purpose: a blended learning approach in English language teaching. *J. Shanghai Jiaotong Univ. (Sci.)* **16**(8), 703–716 (2020)
6. Guo, J., Bai, L., Yu, Z., et al.: An AI-application-oriented in-class teaching evaluation model by using statistical modeling and ensemble learning. *Sensors* **21**(1), 241 (2021)
7. Hu, J.: Teaching evaluation system by use of machine learning and artificial intelligence methods. *Int. J. Emerg. Technol. Learn. (iJET)* **16**(5), 87 (2021)
8. Zhang, X., Shi, W.: Research about the university teaching performance evaluation under the data environment method. *Cogn. Syst. Res.* **56**, 108–115 (2019)
9. Zhu, Y., Lu, H., Qiu, P., et al.: Heterogeneous teaching evaluation network based offline course recommendation with graph learning and tensor factorization. *Neurocomputing* **415**, 84–95 (2020)

10. He, Y., Li, T.: A lightweight CNN model and its application in intelligent practical teaching evaluation. *MATEC Web Conf.* **309**(4), 05016 (2020)
11. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–17 (2018). <https://doi.org/10.1007/s11036-018-1134-8>
12. Liu, S., Bai, W., Liu, G., et al.: Parallel fractal compression method for big video data. *Complexity* **2018**, 2016976 (2018)
13. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. *Mobile Networks and Applications* **26**(5), 1891–1903 (2021). <https://doi.org/10.1007/s11036-020-01725-x>
14. Jinyu, W.: Summative assessment of college English teachers' teaching effect. *IPPTA Q. J. Indian Pulp Pap. Tech. Assoc.* **30**(7), 876–882 (2018)
15. Chai, C., Damnoen, P.S., Phumphongkhochasorn, P.: Theory of planned behavior in support for post Covid-19 new normalization responses of teacherstowards online and blended learning. *Solid State Technol.* **63**(5), 2166–2178 (2020)
16. Graham, C.R., Borup, J., Pulham, E., et al.: K–12 blended teaching readiness: model and instrument development. *J. Res. Technol. Educ.* **51**, 239–258 (2019)
17. Xu, D., Glick, D., Rodriguez, F., et al.: Does blended instruction enhance English language learning in developing countries? Evidence from Mexico. *Br. J. Educ. Technol.* **51**(1), 211–227 (2020)
18. Lontas, J.I., TESOL International Association, Dellicarpini, M.: *The TESOL Encyclopedia of English Language Teaching || Introduction to Assessment and Evaluation*, pp. 1–3 (2018)
19. Peng, R., Fu, R.: The effect of Chinese EFL students' learning motivation on learning outcomes within a blended learning environment. *Australas. J. Educ. Technol.* **37**, 61–74 (2021)
20. Legesse, T., Anbessa, B., Temene, D., et al.: Evaluation of blended fertilizer formulas under limed condition of acid soil on soybean (*Glycine max*) at Asossa District of Benishal-gul Gumuz regional state. *Int. J. Plant Soil Sci.*, 18–29 (2020)
21. Zinovieva, I.S., Artemchuk, V.O., Iatsyshyn, A.V., et al.: The use of MOOCs as additional tools for teaching NoSQL in blended and distance learning mode. *J. Phys. Conf. Ser.* **1946** (1), 012011 (2021). 14pp
22. Thurab-Nkhosi, D., Maharaj, C., Ramadhar, V.: The impact of emergency remote teaching on a blended engineering course: perspectives and implications for the future. *SN Soc. Sci.* **1** (7), 1–19 (2021)
23. Zi, W., Zhou, Y.: Research-based College English blended teaching of fashion major: a case study of Beijing Institute of Fashion Technology. *Int. J. Contemp. Educ.* **3**(2), 55 (2020)





# Design of Golfer's Heart Rate Data Transmission System Based on Machine Learning

Bin-bin Liu<sup>(✉)</sup>

Nanchang Normal University, Nanchang 330032, China  
lbb32900@aliyun.com

**Abstract.** In order to better guarantee the training status and health of sports, a design method of heart rate data transmission system for golf players based on machine learning is proposed. The hardware structure of the system is configured based on the modern background. Based on nRF905 and AT89S52, a set of remote heart rate data transmission system is designed for golfers. Finally, the experiment proves that the performance of the golfer heart rate data transmission system based on machine learning is better in the practical application process. The test results show that the designed system can effectively realize the expected function.

**Keywords:** Machine learning · Golfer · Heart rate data transmission

## 1 Introduction

With the rapid development of science and technology, all kinds of high-tech technology makes the traditional equipment more intelligent. Under the background of Internet of things, the concept of athlete heart rate data transmission system has attracted much attention [1]. Heart rate data transmission system mainly collects golfer information through sensors and other devices, and makes corresponding decisions through data transmission and processing. Its architecture is similar to the Internet of things. This paper uses machine learning technology to build a sensor data transmission system between heart rate data transmission and intelligent terminal and server, in order to solve the problem of insufficient computing, interaction and storage capacity of heart rate data transmission. According to the requirements of data transmission in heart rate data transmission system, the hierarchy of heart rate data transmission system is divided in the system, and machine learning technology is used. The sensor data transmission system of heart rate data transmission is designed.

At present, the common heart rate data transmission system includes non-contact breathing and heart rate signal acquisition system. According to the mechanical characteristics of cardiac ejection contraction process, the piezoelectric ceramic sensor with high sensitivity and good stability is selected to collect the mechanical signal of cardiac shock. The signal is processed by de-noising, filtering and amplification, and the ballistocardiogram (BCG) is obtained by digital acquisition. Secondly, the respiratory signal is extracted by smoothing the cardiac shock, Fast Fourier transform

(FFT) was used to obtain the respiratory signal frequency. Band pass filter was used to remove the respiratory envelope and high-frequency interference of BCG signal, and the J wave peak number per unit time of BCG signal was obtained to calculate the heart rate value. Finally, in order to verify the accuracy and consistency of the system, the respiratory and ECG signals collected by BIOPAC were compared, The results show that the error rate of respiration is less than 4.5%, and the error rate of heart rate is less than 9.7%; There is also a typical system to extract heart rate parameters from face video. In this system, the face video image tracked and recognized by KLT (Kanade Lucas Tomasi) algorithm is converted to YCbCr color space for skin detection, and at the same time, it is converted to CG color channel to extract high-quality photoplethysmograph (PPG) signal, Finally, according to the physiological characteristics of the heart rate signal, the pseudo point noise is removed and the heart rate parameters varying with time are extracted. The system completes the signal acquisition through face changes, the acquisition process is relatively simple, but the accuracy of the acquisition has some defects.

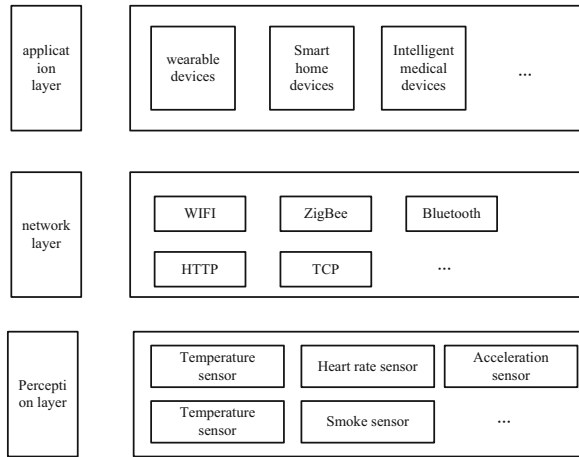
In order to make up for the shortcomings of the above methods, this paper proposes a design of golfer's heart rate data transmission system based on machine learning. In the hardware design of the system, the power consumption requirement of the heart rate data transmission system is studied, and the data transmission of the sensor is realized based on cc2540 chip. The data processing program in the system is added to the software design. According to different sensor data, the algorithm in each level is designed, and the sensor is used to transfer the relevant data to complete the design of the system. The experimental results show that the proposed system can quickly collect the heart rate of golfers.

## **2 Heart Rate Data Transmission System for Golf Players**

### **2.1 System Hardware Design**

The main functions of heart rate data transmission system are data acquisition, data transmission and data processing. Based on these three functions, the heart rate data transmission system is abstracted into hardware module, intelligent module and server module. In the hardware module, the power consumption requirements of the heart rate data transmission system are studied, and the sensor data transmission is realized based on the chip cc2540 [2]. In order to realize the data transmission with cc2540 chip, the system structure is studied, and the sensor data is received and processed. According to the performance requirements of heart rate data transmission system for data transmission and processing, the server module designs data transmission format scheme, server container scheme and a hierarchical software structure. Finally, the information exchange of sensor data among the three modules of the whole heart rate data transmission system is completed [3]. Heart rate data transmission system is usually powered by battery, which has a high requirement for power consumption. Due to the data

transmission system. The network scale of heart rate data transmission system is small, and intelligent terminal equipment is needed as its user interaction system. Therefore, machine learning technology is more applied in heart rate data transmission system [4]. Based on this, the hardware structure of the system is optimized, and the specific structure is shown in the figure below (Fig. 1).



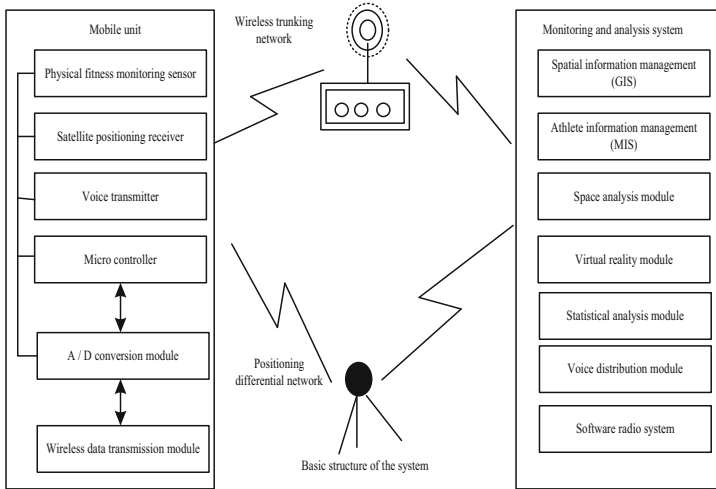
**Fig. 1.** System hardware structure

NRF905 and AT89S52 are selected. NRF905 is a new product developed by Nordic VLSI company in Norway. Before the formal start of the system design work, a series of studies on nRF905 and AT89S52 were carried out, mainly including: the specific structure of nRF905 [5]. The working principle of nRF905, the working principle of AT89S52 and its functions are introduced. The above work is the basis and premise to ensure the smooth development of the system design.

The basic design idea of remote heart rate data transmission system for golf players is as follows: to obtain the heart rate signal of golf players. The equipments needed are ECG electrode and ECG sensor [6]. The heart rate signal is sent by wireless way. At this time, we need to use the wireless signal transponder, which plays a key role is nRF905; to process the heart rate signal and obtain the heart rate data.

When processing heart rate signal, the main method is filtering. After that, the signals need to be stored and processed to facilitate relevant personnel to use the above signals [7]. In this system, the wireless data transmission device is essential. The device mainly includes three parts: transceiver nRF905, AT89S52 MCU and display structure.

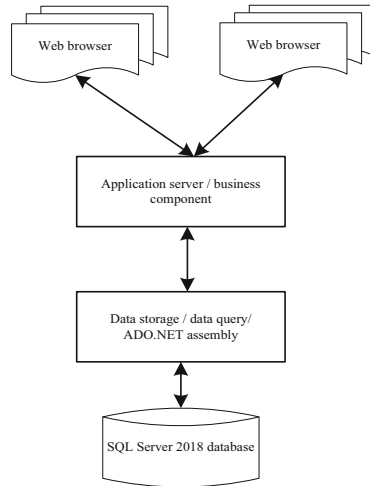
The design of the system, in full consideration of the existing technology, uses physical fitness monitoring sensor, wireless network, a/D conversion and other modules [8]. The specific architecture design is as follows (Fig. 2):



**Fig. 2.** Optimization of functional structure framework of equipment

The wireless network transmission in this system is designed by 1-900/1800 module. The specific design idea is to use communication module to realize communication connection between MCU and computer through mobile communication technology. The specific process is to pack the data into IP data, and put it on the Internet. In the computer room, and send it to the communication module through the base station, then G transmits the IP packet to MCU, and obtains the corresponding processing results through the analysis of the central processor. In the practice design, the embedded 51 MCU system is selected, the remote data transmission adopts mobile network, and the MCU processor adopts the uspd3200 series [9]. It can be regarded as the interface of the whole data acceptance and information development, and the embedded system has good portability.

In this system, three kinds of communication are mainly realized, the first is GPRS and MCU communication, the second is to realize the communication between GPRS communication module and PC; the third is communication between PC and MCU. Based on this, the architecture of communication transmission information management system is optimized, and the specific structure is shown in the following figure (Fig. 3):



**Fig. 3.** Architecture design of communication transmission information management system

The system hardware design mainly includes: heart rate signal acquisition. When collecting the heart rate signal of golf players, there are two key devices, which are ECG electrode and ECG sensor. Among them, the main function of ECG electrode is to collect and process the heart activity signal of golf players, and further transmit it to ECG sensor. The main function of ECG sensor is to process a series of received heart signals, including increasing the original weak signal and filtering the signal. In addition, the ECG sensor can transmit the processed heart rate signal to the wireless signal transponder. The electrocardiograph electrode and electrocardiograph sensor used in this system have the following advantages: it can effectively resist the interference factors in the environment during the working process; the electricity required during the operation is not very much; when collecting the heart rate signal of golf players, the impact on the operators is relatively small, and the golf players can maintain the normal movement state.

The main component of preamplifier circuit is preamplifier, which is AD620 instrument amplifier. The main reason for choosing this component is: control the gain. If the gain is too large, it will cause the ECG amplifier to block, which will affect the normal operation of the system.

Generally speaking, the farthest distance that the heart rate sensor can transmit the heart rate signal is one meter, but in the field of remote monitoring, the monitoring range must exceed one meter, which puts forward new requirements for the transmission of heart rate signal. In order to solve the above problems, nRF905 is selected to meet the needs of remote monitoring with the help of wireless signal transponder. NRF905 also has the ability to receive signals, so in the design of wireless signal receiver, nRF905 is also selected. Generally speaking, in wireless data transmission, there are three types of systems, which are point-to-point system, point to multipoint system and multipoint to multipoint system.

### 2.2 System Software Process Optimization

In order to further optimize the data processing process of the heart rate sensor in the application layer, the main purpose is to increase the data processing program in the system. According to different sensor data, different algorithms can be designed in this layer. After processing, the data can be transferred to the presentation layer in the module or to the server through HTTP protocol. The data presentation layer is the top layer of the intelligent module in the heart rate data transmission system. Users exchange information with heart rate data transmission system through human-computer interaction. The sensor data can be transmitted to users in time, which can make the data produce the maximum value. The interactive mode of the system is introduced into the heart rate data transmission system. In the system, the screen is usually used to display sensor data. However, the information conveyed by simple graphical interface may be ignored by users. In the data presentation module, on the basis of providing a graphical interface to display sensor data, the multimedia technology of the system is designed to convey information. Based on this, the heart rate data acquisition and processing process is optimized as follows (Fig. 4):

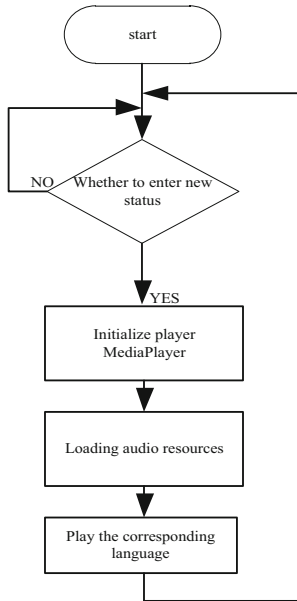


Fig. 4. Heart rate data acquisition flow

The first half of the data receiving process is the same as the sending process. In the second half, when the transceiver is enabled, if a data frame is received, the program will judge whether the data is overflowing, and then judge whether the data frame is too short, whether it is a response frame, if not If the data frame is too short or a response frame, the data frame will be discarded. In the process of building Okumura Hata

model, there will be data transition in the process of training data collection and intelligent data feedback of golfers in big data environment. In order to prevent data transition, it is necessary to modify the calculation process of golfers’ training process. The specific correction methods are as follows:

$$\frac{\delta L_m}{\delta i} = \frac{(k - k_m - N_2)W_0(i)(k - N_i)}{A + w_0(i)\pi} > 0 \tag{1}$$

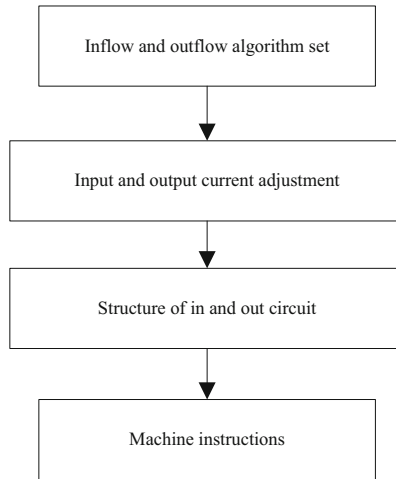
$$\frac{\partial^2 N_2}{\partial i^2} = \frac{(k - k_m - N_2)W_0(i)}{w_0(i)} > 0 \tag{2}$$

In the formula,  $\partial^2 N_2$  is the average weight coefficient of the intelligent customization system for training progress;  $k_m$  is the simulation index of the model data;  $\partial i^2$  is the weight process coefficient of the best operation;  $W_0$  is the expression attribute of big data;  $A$  is the total display amount of the proposed process of the extreme data. In the correction process, the robustness of each module will be reduced, resulting in the increase of operation time. In order to solve this problem, the calculation process needs to be automatically adjusted and optimized. The total amount of the two pole limit data is equivalent to the limit value, and the maximum value and minimum value of the data need to be adjusted and optimized before standard confirmation can be carried out, which can be expressed by formula:

$$G_f = E_p \frac{\delta L_m}{\delta i} \times \frac{\partial^2 N_2}{\partial i^2} \tag{3}$$

In the formula:  $E_p$  represents the maximum critical value of the basic total amount;  $\{h_1, h_2, h_3, \dots, h_p\}$  represents the ordered set that can be collected from the maximum value to the minimum value [10], from which the best optimization deviation of data can be calculated;  $\{X_1, X_2, X_3, \dots, X_p\}$  represents the ordered set composed of loss data in the model; through the above formula, the correction and optimization of the calculation in the intelligent customization of training progress are completed, ensuring the effectiveness of personalized training customization Effectiveness and applicability. Using machine learning technology, the software architecture is designed hierarchically, including view layer, business logic layer, controller layer and Dao layer [11]. MSP430 microcontroller of data acquisition node uses A/D unit to collect data from data acquisition module, which transmits data to CC2420 unit through SPI communication mechanism, and CC2420 transmits data through antenna [12].

In order to improve the efficiency of program transmission, the program is dynamically transmitted according to the language environment and morphemes. The principle of data processing is mainly divided into four stages, as shown in the figure (Fig. 5).



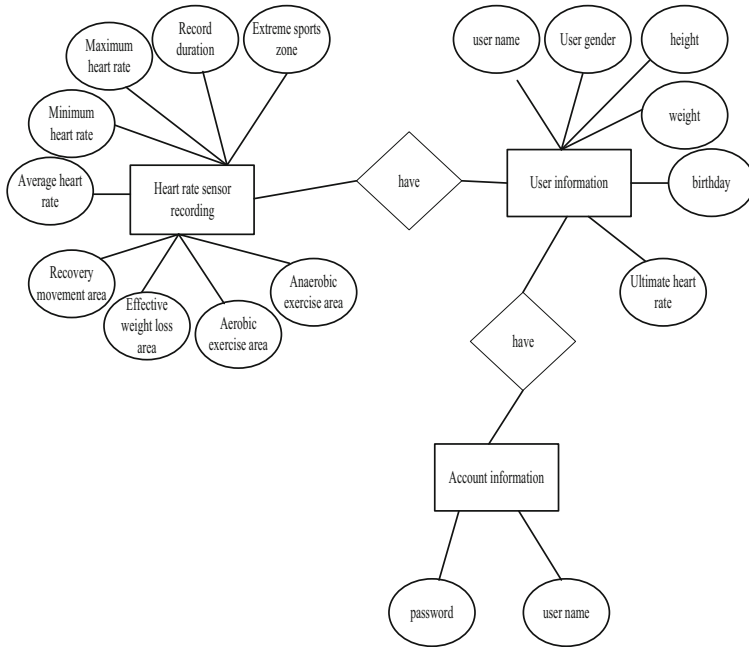
**Fig. 5.** Golfer heart rate processing steps

The program transmission of heart rate collection and transmission of golfers plays an important role in the programming. A layered and open dynamic transmission method is proposed. The method maps the programming language source representing the input and exit to specific technology, which fully reflects the basic operation sequence of electronic access. The input and output algorithm set, current adjustment and circuit construction in the figure are all carried out on the computer transmitter, and the machine instruction is executed on the simulator. The source program is written by advanced language, and can be programmed under complex data structure. The transmission of the input and exit algorithm set is represented by the source program to the middle of the input and exit quantity, and the middle part is encoded by the input and exit quantizer and the operation circuit; the current adjustment maps the source program to the equivalent, only contains one input and exit level, which is usually called the source program Assembler for access. The purpose of current adjustment is to produce high-quality assembly code. The circuit structure is optimized for the calculation of the input and exit with specific attributes.

### 2.3 Realization of Primary Transmission of Golfer's Heart Rate

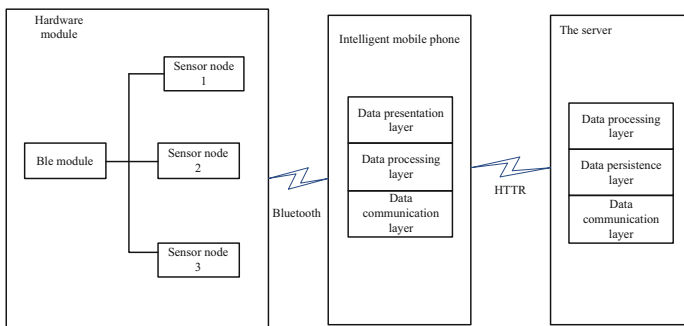
Mysql database is used to store the heart rate sensor data sent by the intelligent terminal. In the heart rate data transmission system, a simple user system is designed and implemented in the server for different users to query personal data. In the design of database table, three entities are designed: sensor record entity, user information entity and account information entity. The sensor record entity has corresponding attributes to store data. The diagram of the three entities is shown below. One user information contains multiple heart rate sensor recording entities and an account information entity (Fig. 6).





**Fig. 6.** Diagram of feature recognition of data to be transmitted

Power consumption is an important index of heart rate data transmission system. The heart rate data transmission system is designed based on low power machine learning technology. The system consists of three parts, including hardware module, intelligent module and server module. As shown in the figure, the three functions of the system can be summarized as data acquisition and transmission, real-time data processing and presentation, and data storage (Fig. 7).



**Fig. 7.** Golfer heart rate data storage model

Because the designed ECG data wireless transmission system can be embedded in the heart rate transmission system, so in practice, the source of ECG data should be the data collected from the athletes' heart electrodes, which needs to be converted into digital signals through a/D conversion. Considering the design of ECG data wireless transmission system hardware, the control module of the system uses 16 bit MSP430 microcontroller, which contains 12 bit a/D, so the collected data is 12 bit, so it uses the 212 format used in MIT-BIH database, that is, three bytes to represent the two collected data, if the two adjacent ECG data collected are 0x3f3 and 0xf3f is represented by three bytes: 0xf3, 0xf3 and 0x3f, that is, the first byte is the lower 8 bits of the first data; the upper 4 bits of the second byte are the higher 4 bits of the second data; the lower 4 bits of the second byte are the higher 4 bits of the first data; the third byte is the lower 8 bits of the second data. These are the two data collected and transmitted in three bytes, as shown in the table below (Table 1).

**Table 1.** ECG data transmission format collected

Byte 1	Byte 2		Byte 3
Data 1 low 8 bits	Data 2 high 4 bits	Data 1 high 4 bits	Data 2 low 8 bits
0, 2, 4	1, 3, 5	6, 7, 8	9, 10

The transmission of psychological data based on the above table transmission format can better guarantee the operation effect of the system, and improve the security rate and timeliness of the system operation.

### 3 Analysis of Experimental Results

In order to verify the practical application effect of the golfer's heart rate data transmission system based on machine learning, the experimental detection is carried out. In order to ensure the experimental detection results, the experimental environment and parameters are set uniformly (Table 2).

**Table 2.** Experimental configuration parameters

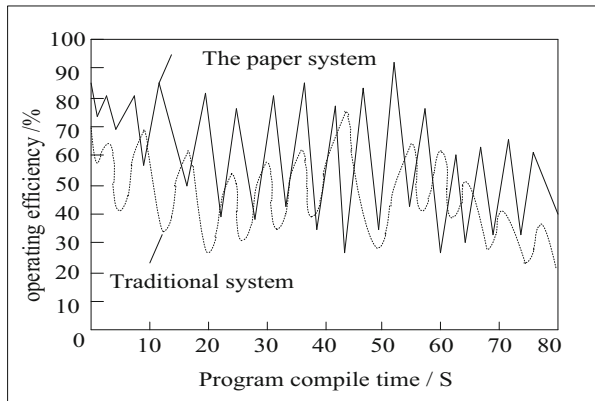
Name	Describe
Libc	System C library
Media Fiamework	Multimedia Library Based on packetvideo
Surface Manager	Display system manager
FreeType	Font engine
SQLite	Lightweight relational database

The selected 20 different golfers were randomly divided into two different groups, the experimental group and the control group. And through statistical software for statistical analysis, through the t test,  $P > 0.05$  (Table 3).

**Table 3.** Basic statistics of athletes

	Average age (years)	Training years (years)	5000m (min)
Experience group	19.9	5.8	16.27
Control group	20.2	6.1	16.14
T test	$P > 0.05$		$P > 0.05$

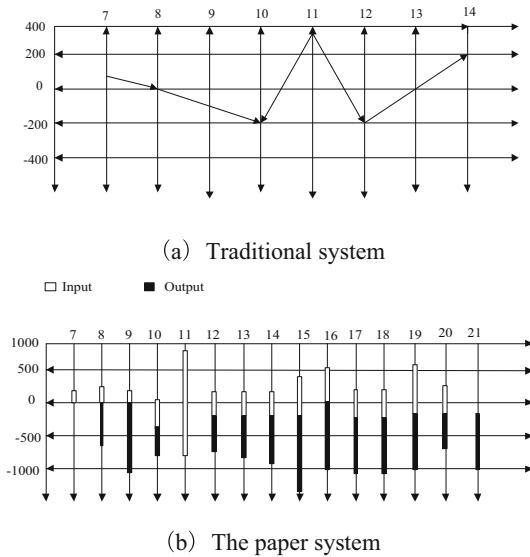
Through the above random selection, and the above-mentioned golfers for a month of training, in order to view the effect analysis of the system in the process of application. Using the electronic program design, the athlete’s in and out can be input by one key, and the athlete’s in and out will be displayed on the record sheet. This paper studies the influence of two kinds of program transmission, file structure and human-computer interface design on the program operation efficiency. In order to verify the efficiency of simultaneous interpreting of two programs in different transmission languages, the traditional program is compared with the program designed for the electronic intake and output (Fig. 8).



**Fig. 8.** Comparison of system operation efficiency

It can be seen from the figure that: the traditional program uses C language to transmit the program, and the running efficiency is in a curve state, while the program designed for electronic in and out volume uses the transmission with specific parameters, and the running efficiency is in a broken line state. The initial operation efficiency is more than 60%, which can meet the requirements of system preparation operation. When the transmission time is 10 s, the efficiency of the traditional program is 10%

higher than that of the electronic access program; when the transmission time is 20 s, the efficiency of the electronic access program is 60% higher than that of the traditional program; when the transmission time is 52 s, the efficiency of the electronic access program reaches the maximum, 92%, while the efficiency of the traditional program reaches the maximum. The efficiency of the program is less than 60%. With the increase of transmission time, the running state of the two programs tends to be stable, but both of them are lower than the initial running efficiency. The final running efficiency of the electronic access program design is 45%, which meets the standard of 30% of the normal running efficiency of the system. Therefore, using specific parameters to transfer the program is more efficient. In order to count the in and out volume of the two programs in different environments, the traditional program is compared with the program designed for psychological transmission volume, and the results are shown in the figure (Fig. 9).



**Fig. 9.** Comparison results of system data transmission

It can be seen from the figure that: the traditional system makes statistics on the data by manual means, and the in and out volume records are relatively rough, showing a broken line, while the in and out volume design program makes statistics on all data in the system memory unit, and the in and out volume records are more detailed, showing a column. According to the record results, the efficiency of the program is compared, and the results are shown in the table (Table 4).

**Table 4.** Running efficiency of two programs under different file structures

Experiment times/time	Traditional procedure	Electronic access program
1	60%	85%
2	55%	90%
3	65%	92%
4	55%	87%
5	72%	91%

It can be seen from the table that: there is a big difference in the operation efficiency of the two programs under the structure, the traditional design program can not accurately obtain the athletes' in and out value, and can only view the general trend, which leads to the decrease of the operation efficiency of the program; while the heart rate data in and out design program can accurately obtain the athletes' in and out value, and the operation efficiency of the program is higher. Therefore, it is more efficient to collect all the data in the system memory and retrieve the key heart rate data in and out according to the basic operation sequence. It is confirmed that the design method of golfer's heart rate data transmission system based on machine learning proposed in this paper has high application value and fully meets the research requirements.

## 4 Conclusion

A remote heart rate data transmission system for golfers based on machine learning is designed. The test results show that the system can achieve the expected function and provide remote heart rate data information. Better training for golf players to improve scientific guidance to ensure the training effect. In the hardware design of the system, the power consumption requirement of the heart rate data transmission system is studied, and the data transmission of the sensor is realized based on cc2540 chip. The data processing program in the system is added to the software design. According to different sensor data, the algorithm in each level is designed, and the sensor is used to transfer the relevant data to complete the design of the system. The experimental results show that the proposed system can quickly collect the heart rate of golfers.

## References

1. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. *Multimedia Tools Appl.* **76**(4), 5787–5802 (2014). <https://doi.org/10.1007/s11042-014-2408-1>
2. Muraro, C., Polato, M., Bortoli, M., et al.: Radical scavenging activity of natural antioxidants and drugs: development of a combined machine learning and quantum chemistry protocol. *J. Chem. Phys.* **153**(11), 114117 (2020)
3. Brugarolas, R., Yuschak, S., Adin, D., et al.: Simultaneous monitoring of canine heart rate and respiratory patterns during scent detection tasks. *IEEE Sens. J.* **19**(4), 1454–1462 (2019)

4. Wang, G., Zhang, S., Dong, S., et al.: Stretchable optical sensing patch system integrated heart rate, pulse oxygen saturation, and sweat pH detection. *IEEE Trans. Biomed. Eng.* **66**(4), 1000–1005 (2019)
5. Yue, H.A., Li, X.B., Cai, K., et al.: Non-contact heart rate detection by combining empirical mode decomposition and permutation entropy under non-cooperative face shake. *Neuro-computing* **392**(5), 142–152 (2020)
6. Gs, A., Mm, B., Lr, B., et al.: A big - data classification tree for decision support system in the detection of dilated cardiomyopathy using heart rate variability. *Procedia Comput. Sci.* **176**(11), 2940–2948 (2020)
7. Chen, X., Pang, L., Guo, P., Sun, X., Xue, Z., Arunkumar, N.: New upper degree of freedom in transmission system based on wireless G-MIMO communication channel. *Clust. Comput.* **22**(2), 4091–4099 (2017). <https://doi.org/10.1007/s10586-017-1513-0>
8. Li, Z., Zi, Y., Chen, J., et al.: Performance-guided maintenance policy and optimization for transmission system of shipborne antenna with multiple components. *Ocean Eng.* **199**(12), 106903–106909 (2020)
9. Yan, Y.: Load characteristic analysis and fatigue reliability prediction of wind turbine gear transmission system. *Int. J. Fatigue* **130**(01), 105259.1–105259.9 (2020)
10. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–10 (2019)
11. Zhang, X.: Design of optical fiber network laser abnormal power data acquisition system based on machine learning. *Laser J.* **41**(9), 111–115 (2020)
12. Zhang, X., Chen, N., Ma, X.: PLC remote data acquisition system based on TCP/IP protocol. *Modul. Mach. Tool Autom. Manuf. Tech.* **15**(4), 91–94 (2020)



# Design of Healthcare Data Analysis System Based on Operational Research and Differential Evolution Algorithm

Xue Jin<sup>1(✉)</sup> and Bin-bin Liu<sup>2</sup>

<sup>1</sup> Hongshan College, Nanjing University of Finance and Economics,  
Nanjing 210003, China  
jx42154@163.com

<sup>2</sup> Nanchang Normal University, Nanchang 330032, China

**Abstract.** Considering the existing medical data has many characteristics, such as many and complex classification, based on operational research and differential evolution algorithm, a medical data analysis system is designed, which consists of distributed storage module, medical intelligent assistant decision module, data cache module, data analysis module, authority verification module, overall decision module and processor module. Among them, data analysis module and overall decision module are the core of the whole system. In the data analysis module, the data output from the data cache module is split and processed by differential evolution algorithm and clustering mining algorithm, and transmitted to the overall decision module. The decision theory in operational research is used to describe and analyze the data analysis process of decision makers, According to the analysis results, the system makes overall decision, and then realizes the design of the medical and health data analysis system. The experimental results show that the system has a high proportion of practical rules extraction, and the time consumption is shorter, which can further improve the level of medical and health data analysis.

**Keywords:** Operational research · Differential evolution algorithm · Health care data · Hadoop system · Data analysis system

## 1 Introduction

In recent years, due to the explosive development of Internet technology, massive and complex user data is generated every moment on the network, and human society has entered the era of big data [1]. Among them, research on big data of medical industry informatization has become the focus at home and abroad. The Research Report of hit2013 in the United States shows that in 2011, the amount of medical data in the United States was about 150eb, and the Caesar group alone stored nearly 32pb of medical data. In 2012, the total amount of medical data in the world has reached 2.7zb, nearly three times that in 2010 [2]. IDC, an international data company, predicts that by 2020, the total amount of global data will be close to 40zb.

In China, the process of medical information construction started late compared with foreign countries. In order to better integrate with the international development trend, “smart medical industry” is listed as a key planning project in the 12th Five Year Plan. According to the statistics of the Ministry of Health, in 2014, China invested 27.51 billion yuan in the informatization construction of the medical industry, and in 2015, the total investment scale exceeded 30 billion yuan. Although the investment is greatly enhanced, the contradiction between the huge medical examination data and the backward medical information storage technology is becoming increasingly significant. According to the latest survey report of Jishi information, under normal circumstances, the amount of medical information data for a healthy young person is basically within 1MB, for middle-aged people with some physical problems it is about 40MB, and for the middle-aged and elderly people with several chronic diseases, it will reach 3-5gb. Due to the limitation of technical means, the vast majority of medical institutions and hospitals adopt the traditional as.

The system of literature [3] was first proposed by renowned American scholar Morris Colen in the 1990 s, which belonged to the typical C/S architecture, the client/server working mode. In the system, the clients distributed in various departments are connected with the server of the data center through the enterprise office network. Under normal circumstances, SQL server or Oracle database is installed on the server of data center to store medical data, and the electronic medical records generated by various medical examination reports and medical diagnosis results of patients, as well as various medical audio and image data are stored in the database. The client system submits the query, insert or modify instructions to the server’s medical database through UI, and the instruction execution results are fed back to users at all levels through UI. The advantages of his are simple structure, mature technology and relatively easy system design and implementation.

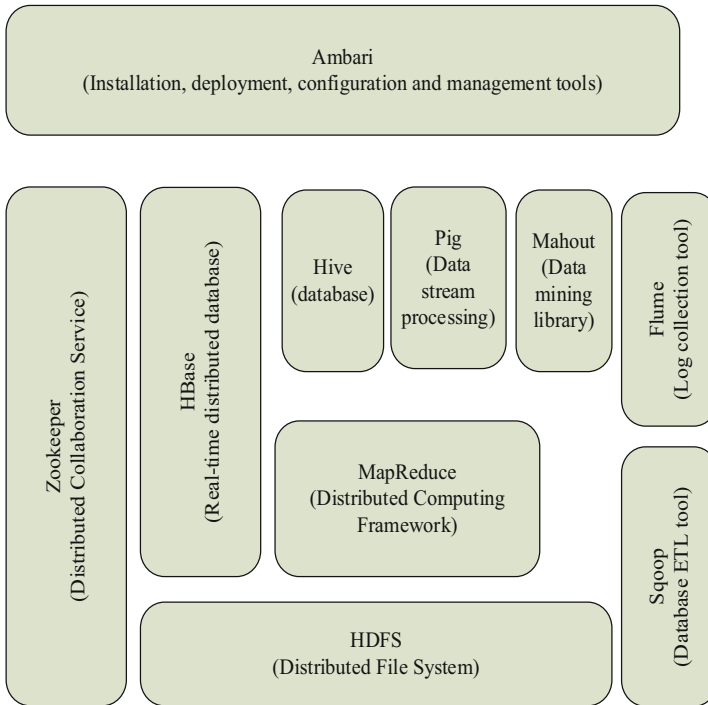
However, when dealing with large-scale data sets, due to the limitation of processing capacity, this paper proposes a healthcare data analysis system based on operational research and differential evolution algorithm to solve the problem that the traditional relational database can not meet the needs of deep mining and big data analysis.

## **2 Design of Healthcare Data Analysis System Based on Operational Research and Differential Evolution Algorithm**

### **2.1 Design Distributed Storage Module**

In the distributed storage module, in order to realize the classified storage of heterogeneous data in the traditional medical database, it needs the help of Hadoop system component collection and third-party integration tools [4–6]. The component system of Hadoop project is shown in Fig. 1.





**Fig. 1.** Hadoop component systems for the project

During the design of distributed storage module, hive database, HBase database and sqoop tool are used in the component system.

- (1) Hive is a database based on HDFS designed by Facebook. Its feature is that it can convert database files into data forms, and all SQL commands will be converted into map/reduce parallel processing program, which enables users to complete data query, writing and other related operations in the form of commands without programming.
- (2) HBase is a real-time distributed database based on “column mode” running on the top of HDFS. Different from the traditional relational database based on parallel schema, HBase uses BigTable data model to store data in the form of table, and divides the table into rows and columns. The sparse permutation mapping table (key/value) composed of row keywords and column keywords can be used for map/reduce processing. Compared with hive, HBase has the function of random read-write and real-time access to large-scale data, and BigTable’s data model perfectly combines data storage and parallel computing, which is very suitable for unstructured data storage and management.
- (3) Sqoop tool is an independent open source sub project developed by Apache project team, and it is the first third-party integration tool of Hadoop to realize data transmission between traditional database (such as SQL server or Oracle) and Hadoop.

Distributed storage module is the key part of the system, and it is also the basis of medical intelligent decision support and medical big data analysis.

Before distributed storage and processing of original healthcare data, cellular and HBase are installed on named nodes, and then use the Java API provided by sqoop tool to connect with the traditional medical database. Then, for all kinds of data that need to be imported, the structure properties are judged.

- (1) If it is structured data, sqoop tool will connect hive through JDBC/ODBC interface, and then query whether the storage form corresponding to the data already exists, if not, create a new table and save it in hive; If it already exists, judge whether the amount of data exceeds the set threshold. If not, store it directly in a hive; If it exceeds, you need to add a partition before saving it in the hive.
- (2) When the data is unstructured, sqoop will connect to HBase through the HBase interface and submit the insertion request; After the request is responded, scan the HBase table and locate the insertion position. At the same time, set the timestamp to insert the data into the HBase database [7–9].

In order to write data correctly, it is necessary to configure Hadoop cluster, including the running parameters of the guard and the running environment of Hadoop. The corresponding relationship between configuration items and configuration files is shown in Tables 1 and 2.

**Table 1.** Daemon and configuration items

Serial number	Daemon name	Corresponding configuration item
1	Name node	HADOOP_NAMENODE_OPTS
2	Data node	HADOOP_DATANODE_OPTS
3	Job tracker	HADOOP_JOBTRACKER_OPTS
4	Task tracker	HADOOP_TASKTRACKER_OPTS

Each daemons runs independently in the background of Linux system. The namenode process is responsible for namespace management and file access, the datanode process is responsible for connecting data nodes, and the jobtracker and tasktracker are responsible for user job scheduling and slicing execution.

**Table 2.** Hadoop main configuration files for clusters and related functional descriptions

Serial number	Profile name	File function description
1	/etc./host	Specify the corresponding relationship between name node, datanode and IP address
2	/etc./profile	System environment variable configuration file
3	/Hadoop/conf/hadoop-env.sh	Configure Java home and Hadoop environment variables

(continued)

**Table 2.** (continued)

Serial number	Profile name	File function description
4	/Hadoop/conf/core-site.xml	Configure HDFS temporary directory, address and port number
5	/Hadoop/conf/hdfs-site.xml	Configure the log file directory and the number of data backups
6	/Hadoop/conf/mapred-site.xml	Configure the IP address and port of job tracker
7	/Hadoop/conf/master	Add named node IP
8	/Hadoop/conf /slave	Add data node IP
9	/Hive/conf / hive-site.xml	Add HDFS root directory, jobtracker IP and port
10	/HBase/conf/hbase-site.xml	Add master node IP, port and zookeeper location directory
11	/Sqoop/conf sqoop-env.sh	Add HDFS directory, hive and HBase location directory

After the Hadoop cluster configuration is completed and the original data is imported successfully, the client starts to create the distributed file system HDFS. The data writing process is as follows:

- (1) The data layer client development library starts the data node and sends RPC connection access request to the named node of the access control layer.
- (2) The named node checks whether the file to be created already exists and the operation permission of the creator. If the check is successful, a record is created for the file; If the check fails, an exception will be thrown to the client [10, 11].
- (3) When the RPC write request is responded, the client development library divides the files to be written into multiple packets, applies for new blocks from the named node, and generates a “block report” from the mapping list of local files and HDFS data blocks to submit to the named node.
- (4) The named node returns the configuration information of the managed data node to the client, and the client will write it to each data node in the form of pipeline according to the IP address of the data node.

In the process of data writing, the data in HBase database can be imported into HDFS through its own writing tool “hbase org.apache.hadoop.hbase.mapreduce.Export”; The data in hive database can be imported into HDFS by executing the “insert rewrite” command. The rest of the implementation process is completed by the bottom layer of Hadoop without user intervention.

## 2.2 Design of Medical Intelligent Auxiliary Decision Making Module

In the process of patients' actual medical treatment, a large number of medical examinations are usually needed. However, due to the difference in patients' physique, the medical examination items for the same type of diseases may show different data results in the examination process of different patients. Therefore, after the medical examination, some patients need to undergo a period of observation and treatment to determine the specific types of the disease. The research purpose of medical intelligent decision-making module is that doctors can quickly identify patients' condition and get more accurate pre diagnosis conclusions by means of advanced IT technology and supported by medical and health data analysis system, so as to optimize clinical process and improve the efficiency of hospital [12–14].

In the existing electronic medical records of hospitals, there are many medical examination data and personal information of patients with confirmed diseases. Many hospitals are limited by IT technology level, and effectively extract and make full use of many practical information in the electronic medical records. The intelligent decision-making module can map/reduce all the electronic medical records stored in HDFS, summarize the data values of various medical examination items of different diseases and generate auxiliary detection template for doctors' reference. The study can reduce the rate of misdiagnosis to some extent, and provide valuable advice for doctors.

The process of map/reduce processing is as follows:

Mapper algorithm:

- (1) Read EMR file, if file  $\neq \Phi$  & Not EOF (all files) then loop to read the string into the variable STR;
- (2) If STR = diagnosis result then key = disease name;  
Else continues to read backward;  
End If
- (3) If STR = medical examination data then key = disease name & medical examination item name;  
Value = the value of a medical examination item corresponding to the disease;  
Else continues to read backward;  
End If
- (4) Write each group (key, value) key value prior to the intermediate result file;

Because a certain disease may correspond to multiple medical examinations, the mapper algorithm adopts the way of key value combination in the design process, which generates multiple (key, value) key value pairs for each medical examination result of the disease and outputs them to the intermediate result file for subsequent processing by the reducer algorithm.

Reduce algorithm:

(1) If intermediate result file  $\Phi$  & Not EOF (all intermediate result files) then circularly reads each group of (key, value) key value pairs;

(2) When a key value is read for the first time

If key = name of a specific disease & name of a specific medical examination item then

Max= Value; Min= Value

(3) Cycle to read all (key, value) key value pairs, for the same key value;

If Value' > Max then Max = Value';

If Value' < Min then Mim = Value';

(4) The (min, max) interval corresponding to each key value is written into the final result file; The innovation of the algorithm is that the mapper process adopts the variable way of key value combination, and transforms the merging result from data statistics to interval generation, which solves the problem of different forms of key values of the same nature, and makes the merging result get good extension and expansion, which is more practical.

After mapping/reduce processing of all electronic medical records in the hospital, all medical examination items corresponding to each type of disease form a reliable interval value. However, there is no effective data association between the independent interval result files, and many similar diseases may be accompanied by similar medical examination results. Therefore, in order to achieve the accurate characterization of the disease, it is necessary to further mine and analyze all kinds of data.

The coupling degree model is applied to the generation of association rules, and combined with Apriori algorithm, the optimization algorithm is described as follows:

Algorithm 1: the process of frequent itemsets discovery algorithm is as follows:

Input: initial transaction set B, and set the minimum support count threshold min<sub>supp</sub>

Output: frequent item set L

L1 =fmd\_frequent\_1-item sets(B); //Get 1 itemset

For(k=2; Lk -1 != null; k++) do{

Ck=apriori\_gen(Lk-1); //Generating candidate sets

For each t in B do {

Ct=subset(Ck, t);

For each c ∈ C do

C.Count ++ ;

}

Lk={ c ∈ Ck | C.Coun t ≥ min<sub>supp</sub> }

}

Return L; //Return frequent itemsets

After generating frequent projects, a subroutine is designed to generate candidate sets\_Rule() and pruning subroutine calculate\_X(), which is the auxiliary association rule algorithm.

(1) generate\_Rules subroutine

The function of this subroutine is to generate a candidate k-item set according to the (k-1) item frequent set in Apriori algorithm, and modify the steps of  $\Pi [1] < \Pi [2] < \dots < \Pi [k]$  in the connection process of the original algorithm to: compare the current two sets  $I_m$  and  $i_n$ , if there is  $(I_m [1] = i_n [1]) \wedge (I_m [2] = i_n [2]) \wedge \dots \wedge (I_m [k-2] = i_n [k-2])$  and the last one satisfies  $(I_m [k-1] = i_n [k-1])$ , Then  $I_m$  and  $i_n$  are connected to produce k-item sets.

(2) calculate\_X subroutine

The function of the subroutine is to introduce the mathematical model calculation into the pruning process of the original algorithm, prune and delete the new error strong association rules in the k-term set, and reduce the time and space cost of the algorithm [11, 12].

Algorithm 2: the implementation process of association rule generation algorithm is as follows:

Input: frequent item set L, set the minimum confidence threshold  $\min\_Conf$ , minimum coupling threshold  $\min\_Cou$

Output: association rule set

```

For(m=2; m++) do{
Hm = {rule file of Lk has m items};
Generate rules by Lk and Hm;
Generate Hm+1 by Hm using Apriori algorithm;
Calculate confidence conf // calculate confidence
If conf < min_conf then
Delete hm+1;
Else{
Calculate cou by EQ (3.10) // calculate coupling degree
If |Cou| > min_Cou then
Output rule (Lk→hm) to Hm+1;
End if
Generate rules by Lk and Hm+1;
End if
}
Return rules; // Return to build rule

```

The innovation of the algorithm lies in: introducing the coupling degree model, and dividing the positive and negative confidence interval through the calculation of the coupling degree value, which effectively inhibits the generation of wrong association rules in the interest degree model.

### 2.3 Design Data Writing Interface Module

The data writing interface module provides the data writing interface. The data mainly comes from the data source provided by the data provider. The data source needs to complete the data cleaning and the corresponding data formatting process. The data writing interface module mainly has the following sub functions: identity verification function and data submission function. Before the data provider writes the information to the database, it needs to submit the identity information to the module. The module then sends the information verification request to the permission verification module. The permission verification module will return the consent or refusal according to the submitted identity information. When the application is submitted, the module helps the data provider to complete the data submission.

### 2.4 Design Data Cache Module

The data writing interface module provides the data writing interface. The data mainly comes from the data source provided by the data provider. The data source needs to complete the data cleaning and the corresponding data formatting process. The data writing interface module mainly has the following sub functions: identity verification function and data submission function. Before the data provider writes the information to the database, it needs to submit the identity information to the module. The module then sends the information verification request to the permission verification module. The permission verification module will return the consent or refusal according to the submitted identity information. When the application is submitted, the module helps the data provider to complete the data submission.

### 2.5 Design Data Analysis Module

Data analysis module is designed based on differential evolution algorithm. The main function of data analysis module is to realize the integration of differential evolution algorithm and clustering mining algorithm. The module includes the realization of privacy protection in data analysis and the corresponding clustering algorithm.

The module can be divided into two sub modules: differential privacy mechanism module and clustering algorithm module. The main function of the differential privacy protection mechanism module is to query the corresponding data in the database according to the query function provided by the differential evolution algorithm. The process is mainly divided into two steps. First, read the value of the privacy operation from the database table, then determine the size of the local sensitivity according to the data and the query function, and calculate the noise size through the privacy operation and the local sensitivity, and the real query results and the noise are added to return the query results with noise. The clustering algorithm module mainly reads the data in the cache database, reads the center point from the database center point data table, marks the data after calculation, and stores the data in the final database after marking. When a batch of cache data is clustered, the clustering module will further calculate the

clustered data. In this step, the cluster distance and intra cluster distance of each cluster will be calculated, and the inter cluster similarity and inter cluster similarity will be generated, and then the corresponding clusters will be split or combined after comparing with the threshold.

## 2.6 Design Permission Checking Module

The authority verification module mainly controls the data writing and data display. The related functional modules mainly include data interface module and data display module. The verification module mainly verifies the content of the data provider's identity information provided by the data writing interface. If the data provider is in the trusted list, it is allowed to write to the database, otherwise, it is refused to submit the application. The function related to the data display module is to verify the validity of the identity of the data miner according to the login information requested by the data miner. If it passes the verification, it will be allowed to log in to the system and display the corresponding data, otherwise it will be refused the login application [15, 16]. □

## 2.7 Design the Overall Decision-Making Module

To design the whole decision-making module based on operations research is to make the whole decision of the system through the decision theory of operations research. Decision making is a kind of activity that people often encounter in production and life [17]. The purpose of decision-making is to summarize the basic rules of decision-making activities, formulate the rules and methods that should be followed when making decisions, so that in the problem of multiple options selection, the decision-maker can make the optimal choice, in order to obtain the best decision-making results or meet the expected purpose of decision-making. According to their nature, decision-making problems can be divided into three types: deterministic decision-making, risky decision-making and uncertain decision-making. Deterministic decision-making refers to the decision-making when the decision-making environment is certain and the decision-making result is also certain. In the deterministic decision-making problem, the desired goal and initial state are generally determined [18, 19]. At the same time, there are many alternative decision-making schemes, and the profit values of these decision-making schemes are determined and can be calculated. For this kind of deterministic decision-making problem, mathematical programming method can be used to solve the optimal decision-making scheme. Risk based decision-making refers to the decision-making when the decision-making environment and decision-making results are uncertain, but their probability of occurrence is known. When making risk decision, the expected value is usually regarded as the decision criterion. The most commonly used decision criteria are maximum expected return and minimum opportunity loss. Uncertain decision-making refers to decision-making under uncertain decision-making environment and unknown decision-making results [20–23]. In uncertain decision-making problems, decision-makers make decisions according to their own subjective attitude. The principles to solve uncertain decision-making



problems usually include pessimistic decision-making criteria, optimistic decision-making criteria, possibility criteria, minimum opportunity loss criteria and eclecticism criteria. The pessimistic decision criterion is to find the minimum profit value of each decision scheme first, and then find the maximum value of these profit values, so as to determine the decision scheme; The optimistic decision criterion is to find the maximum profit value of each decision scheme, and then determine the maximum value of these profit values, so as to determine the equal possibility criterion of decision scheme. It means that when making a decision, if it is uncertain which state has a higher probability and which state has a lower probability, it is assumed that their probabilities are equal. Based on this, the expected value of each strategy profit is calculated, and then the maximum value is found from these expected values to determine the decision scheme; The minimum opportunity loss criterion is also called the minimum regret value criterion. It first converts each element of the return matrix into the corresponding opportunity loss value, that is, the loss caused when the decision scheme with the largest return is not selected, and then determines the maximum opportunity loss value of each decision. Then the minimum value is found from all the maximum opportunity loss values, and the decision scheme is determined. The principle of eclecticism is to first set an optimistic coefficient, then use the optimistic coefficient to get the function value of the maximum and minimum return value of each strategy, and then determine the decision strategy.

## 2.8 Design Processor Module

The processor model selected for the processor module is gt436, and the specific technical data is shown in Table 3.

**Table 3.** GT436 Specific technical data

Serial number	Project	Specific data
1	Instruction cache	16 KB
2	Controller	LCD
3	Data cache	16 KB
4	Guidance system	NAND flash
5	UART	3 Channels
6	PWM timer	4 Channels
7	I/O port	
8	RTC	
9	10 bit ADC	8 Channels
10	SPI	2 Channels
11	Bus-iic interface	
12	USB device	
13	USB	
14	MMC card interface & SD main card	
15	PLL frequency multiplier clock	

### 3 System Performance Testing

#### 3.1 Experimental Environment

According to the actual operation of the system, the experiment is carried out in the network center of Nanhua Hospital Affiliated to Nanhua University. Combining with the existing test system database and using Hadoop cluster to realize data classification and replication, the experiment is imported into HDFS, which lays the foundation for the subsequent intelligent decision support test and big data statistical analysis test.

#### 3.2 Experimental Equipment and Topology

The main hardware equipment used in the experiment includes: hospital internal medical server, personal PC, optical switch and disk array storage platform. The equipment list is shown in Table 4.

**Table 4.** List of hardware test equipment

Equipment name	Number	Equipment model	Parameter configuration		
			CUP	Memory	Hard disk
medical server					
Personal PC	1	Lenovo x3850	Intel(R) Xeno(R) E7-4820V3 @ 1.9 GHz	256 GB	5 TB
Disk array	3	Lenovo Qitian M4350	Intel (R) Core (TM) i5-3470 @ 3.2 GHz	8 GB	1 TB
Storage platform	1	EMC vnx 5300	Maximum number of drives: 125; Known number of virtual resource blocks: 8 total capacity: 10 TB		
Optical switches	1	EMC DS-300B	Interface/Quantity: optical fiber/24 ports; Throughput: 8 gb/s; Fuselage structure: 1RU		

The system is developed with Java language tools, and the built Hadoop cluster runs on the Linux operating system platform. The list of experimental software equipment is shown in Table 5.

**Table 5.** List of software laboratory equipment

Serial number	Equipment name	Version number
1	Development tool	JDK 1.7
2	Operating system	Ubuntu Linux 14
3	Distributed cluster	Hadoop-2.6.28

All the hardware devices used in the experiment are deployed in the central computer room of Nanhua Hospital Affiliated to Nanhua University. The EMC vnx

5300 disk array storage platform is used to realize the storage expansion of his medical server, and the devices are interconnected through the EMC ds-300b optical switch.

The experiment uses three Ubuntu Linux nodes to build a distributed cluster, including a name node and two data nodes. Hadoop distributed architecture software is installed respectively. User interface and operation center are all encapsulated in name node, hive and HBase databases and data export tool scoop are installed, and all user operations will be completed in name node.

### 3.3 Cluster Building

Due to the limitation of experimental conditions, it is impossible to import all the existing medical data into HDFS during the implementation of distributed storage of medical data. Therefore, 50535 patient electronic medical records were collected as the test data set.

Before the construction of Hadoop distributed cluster, the host name and IP address of each node (master node and slave node) in the cluster should be set to ensure that the host name and IP address can form a correct mapping relationship after the cluster is running and can be resolved normally. The configuration information of platform nodes is shown in Table 6.

**Table 6.** Hadoop platform node configuration information

Name of node	Node type	Host name	IP address	Gateway
Name	Master	Master	192.168.1.14	192.168.1
Node	–	Hadoop	1	1
Data Node 1	Slave	Slave1. Hadoop	192.168.1.142	192.168.1.1
Data Node 2	Slave	Slave2. Hadoop	192.168.1.137	192.168.1.1

After the information of each node is set up, the Hadoop distributed experimental platform should be built according to the actual information in its configuration table. The main steps of cluster implementation are as follows:

- (1) Using VI text editor, modify / etc. / host name file to specify host name and IP address for each node; Then modify the / etc. / host file, configure the host DNS server information, and add the host names and IP addresses of all nodes in the cluster.
- (2) In the actual running process of Hadoop, name node will manage (start or stop) the daemons of each datanode through SSH (secure shell). When executing instructions between nodes, there is no password input support, so it is necessary to set SSH free password access for each node to log in to each other.  
In order to realize public key authentication without password, a key pair, including a public key and a private key, should be generated on the server master, and then the public key should be copied to all the client slave. When

master connects to salt through SSH, salt will generate a random number, encrypt the random number with master's public key, and then send it to master; After receiving the encrypted data, the master uses the private key to decrypt, and then sends the decrypted data back to slave; After slave confirms that the decryption number is correct, master is allowed to connect with it.

The generated key pair is stored in the/home/Hadoop/. SSH directory, and then `cat ~/.ssh/ID` is executed `_rsa.pub >> ~/.ssh/authorized_Keys`, which sets the `ID_ Rsa.pub` is added to the authorized key, and then the SSH copy ID command is used to transfer the public key to the slave node.

All relevant settings need to be completed after the SSH service is restarted. The above steps are copied to other nodes to realize SSH password free access to all nodes in the cluster.

- (3) Install JDK 1.7, edit / etc. / profile file, and configure Java environment variables. The SCP command is used on the master node to copy all files after Java installation to all slave nodes, and set the same environment variables to ensure that the cluster can run Java programs normally.
- (4) After installing Hadoop, add the location of Hadoop directory to / etc. / profile file, edit `hadoop-env.sh` file and add Java home directory. In the process of Hadoop configuration, `core-site.xml`, `hdfs-site.xml` and `mapred-site.xml` are mainly involved, which correspond to HDFS distributed file system and map / reduce processing architecture respectively.

When configuring Hadoop on the master, you need to add the address and port number of name node in the `core-site.xml` file first; Then, the backup number of data is configured in `hdfs-site.xml` file; Finally, set the job tracker address and port number in the `mapred-site.xml` file. After copying all the configurations to slave node, formatting HDFS file system and closing firewall, the cluster can be built.

The working state of the built distributed cluster is shown in Fig. 2.

```
Datanodes available: 2 (2 total, 0 dead)
Name: 192.168.1.137:50010
Decommission Status : Normal
Configured Capacity: 16651051008 (15.51 GB)
DFS Used: 28687 (28.01 KB)
Non DFS Used: 5241806833 (4.88 GB)
DFS Remaining: 11409215488(10.63 GB)
DFS Used%: 0%
DFS Remaining%: 68.52%
```

```
Name: 192.168.1.142:50010
Decommission Status : Normal
Configured Capacity: 8039419904 (7.49 GB)
DFS Used: 28672 (28 KB)
Non DFS Used: 3985838080 (3.71 GB)
DFS Remaining: 4053553152(3.78 GB)
DFS Used%: 0%
DFS Remaining%: 50.42%
```

**Fig. 2.** Working state of distributed clusters

After the successful construction of the distributed cluster, sqoop tool and JDBC interface program developed on the Java language platform are used to connect with the SQL Server 2012 database of the medical server and import the data to the master node.

### 3.4 Practical Rule Extraction Test

After the test data set is successfully imported, a test template is generated. The result interval of each disease check item in the template is saved in the Reduce Out Put output file. In order to further test the practical rule extraction performance of the system, part of the data in the training data set accideiits.data, which is commonly used in the field of data mining, is selected as the test data source, and extracted 500 rules for testing, of which 320 rules are defined as practical strong association rules, and the rest are non-practical rules. The test results are shown in Table 7.

**Table 7.** Test results

Threshold setting	min_Cou = 0.3			min_Cou = 0.4		
	Dig total (Article)	Practical rules (Article)	Practical proportion	Dig total (Article)	Practical rules (Article)	Practical proportion
min_conf = 0.7 min_supp = 0.1	384	315	82%	293	226	77%
min_conf = 0.7 min_supp = 0.2	202	149	74%	165	104	63%
min_conf = 0.7 min_supp = 0.3	97	66	68%	82	48	59%

It can be seen from the above results that when the minimum confidence (min\_conf) threshold is the same, setting a smaller minimum support (min\_supp) threshold can mine more association rules, and the proportion of practical strong association rules is also higher. high. Under the same minimum confidence and minimum support settings, the system has a high proportion of practical rule extraction, thus effectively completing the pruning of useless association rules.

### 3.5 Time Consumption Test

Table 8 shows the time-consuming changes of the design system when the number of electronic medical records continues to grow.

**Table 8.** Time consumption test results

Serial number	Number of electronic medical records (copies)	Time consuming (ms)
1	10000	103
2	20000	148
3	30000	169
4	40000	187
5	50000	206

The experimental results show that, with the increasing number of electronic medical records, the consumption time of the design system grows slowly, and the total consumption time is short, mainly because the system uses differential evolution algorithm to optimize the performance of the system transmission module.

## 4 Conclusion

With the rapid development of medical industry information construction, all kinds of heterogeneous data show an explosive growth trend. However, the traditional systems have encountered varying degrees of technical bottlenecks, so it is difficult to achieve the two-way expansion of hardware performance and work efficiency. Therefore, a healthcare data analysis system based on operational research and differential evolution algorithm has been studied and implemented, which makes up for the shortcomings of the existing system to a certain extent. However, in the process of this study, only the sample data was analyzed, which led to the results having a certain impact, and the next step will continue to be analyzed.

## References

1. Domadiya, N., Rao, U.P.: Improving healthcare services using source anonymous scheme with privacy preserving distributed healthcare data collection and mining[J]. *Computing* **103** (4), 1–23 (2021)
2. Shilo, S., Rossman, H., Segal, E.: Axes of a revolution: challenges and promises of big data in healthcare. *Nat. Med.* **26**(1), 29–38 (2020)
3. Stahlhut, R.W., Porterfield, D.S., Grande, D.R., et al.: Characteristics of population health physicians and the needs of healthcare organizations. *Am. J. Prev. Med.* **60**(2), 198–204 (2021)
4. Fan, K., Zhu, S., Zhang, K., et al.: A lightweight authentication scheme for cloud-based RFID healthcare systems. *IEEE Netw.* **33**(2), 44–49 (2019)
5. Mbizvo, G.K., Bennett, K.H., Schnier, C., et al.: The accuracy of using administrative healthcare data to identify epilepsy cases: a systematic review of validation studies. *Epilepsia* **61**(5), 1–17 (2020)
6. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–17 (2018). <https://doi.org/10.1007/s11036-018-1134-8>

7. Xiao, M., Hill, C., Vacquier, M., et al.: Retrospective analysis of the effect of postdischarge telephone calls by hospitalists on improvement of patient satisfaction and readmission rates. *South. Med. J.* **112**(7), 357–362 (2019)
8. Johansson, M., Finizia, C., Persson, J., et al.: Cost-effectiveness analysis of voice rehabilitation for patients with laryngeal cancer: a randomized controlled study. *Support. Care Cancer* **28**(11), 5203–5211 (2020)
9. Hassan Zadeh, A., Zolbanin, H.M., Sharda, R., et al.: Social media for nowcasting flu activity: spatio-temporal big data analysis. *Inf. Syst. Front.* **21**(4), 743–760 (2019)
10. Liu, S., Bai, W., Liu, G., et al.: Parallel fractal compression method for big video data. *Complexity* **2018**, 2016976 (2018)
11. Golob Jr., J.F., Kreiner, A.: Prevention of surgical infections: building or renovating a new intensive care unit. *Surg. Infect.* **20**(2), 107–110 (2019)
12. Zhang, Y., Rodrigues, J., Seah, W., et al.: Guest editorial special issue on wearable sensor-based big data analysis for smart health. *IEEE Internet Things J.* **6**(2), 1293–1297 (2019)
13. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. *Multimedia Tools Appl.* **76**(4), 5787–5802 (2014). <https://doi.org/10.1007/s11042-014-2408-1>
14. Wang, Q., Liu, R., Men, C., et al.: Temporal-spatial analysis of water environmental capacity based on the couple of SWAT model and differential evolution algorithm. *J. Hydrol.* **569**, 155–166 (2019)
15. Pan, Z., Fang, S., Wang, H.: LightGBM technique and differential evolution algorithm-based multi-objective optimization design of DS-APMM. *IEEE Trans. Energy Convers.* **36**, 441–455 (2020)
16. Hua, Y., Sui, X., Zhou, S., et al.: A novel method of global optimisation for wavefront shaping based on the differential evolution algorithm. *Opt. Commun.* **481**, 126541 (2021)
17. Elaziz, M.A., Xiong, S., Jayasena, K., et al.: Task scheduling in cloud computing based on hybrid moth search algorithm and differential evolution. *Knowl.-Based Syst.* **169**(APR.1), 39–52 (2019)
18. Poczeta, K., Kukasz, Ł., et al.: Reprint of: analysis of an evolutionary algorithm for complex fuzzy cognitive map learning based on graph theory metrics and output concepts. *Biosystems* **186**, 104068–104068 (2019)
19. Ye, X., Chen, H., Kuang, Q., et al.: Solving time-dependent reliability-based design optimization by adaptive differential evolution algorithm and time-dependent polynomial chaos expansions (ADE-T-PCE). *Microelectron. Reliab.* **114**(3), 113815 (2020)
20. Dashti, A., Noushabadi, A.S., Raji, M., et al.: Estimation of biomass higher heating value (HHV) based on the proximate analysis: smart modeling and correlation. *Fuel* **257**(Dec.1), 115931.1–115931.11 (2019)
21. Ma, X., Zhang, K., Zhang, L., et al.: Data-driven niching differential evolution with adaptive parameters control for history matching and uncertainty quantification. *SPE J.* 1–18 (2021)
22. Yza, B., Hwa, C., Qla, B., et al.: Automatic data clustering using nature-inspired symbiotic organism search algorithm - ScienceDirect. *Knowl.-Based Syst.* **163**, 546–557 (2019)
23. Zhang, K., Zhu, G., Ma, J., et al.: Parameter analysis and estimates for the MODIS evapotranspiration algorithm and multiscale verification. *Water Resour. Res.* **55**(3), 2211–2231 (2019)



# Research on Reliability Evaluation of Intelligent Energy Integrated Service Platform Based on Hierarchical Linear Model

Wen-Lin Xu<sup>1</sup>(✉), Xiao-xiao Liu<sup>2</sup>, Chao Li<sup>2</sup>, Zi-peng Hu<sup>2</sup>,  
and Kai Liu<sup>2</sup>

<sup>1</sup> Marketing Department, State Grid Hunan Electric Power Co., Ltd.,  
Changsha 410000, China  
dsf2541521@163.com

<sup>2</sup> State Grid Hunan Integrated Energy Service Co., Ltd.,  
Changsha 410000, China

**Abstract.** Intellectualization is a new form of energy services, which can accelerate the integration of different energy services, expand the scope of traditional energy services, and meet the various service requirements of users. In order to improve the use effect of the platform, the reliability evaluation method of the intelligent energy integrated service platform based on hierarchical linear model is studied. The hierarchical linear model is used for data mining and prediction of the intelligent energy integrated service platform, and the functional structure of the platform is optimized, so as to better improve the security of the platform and the utilization efficiency of the intelligent energy integrated service platform, so as to better meet the growing demand of smart energy services of users.

**Keyword:** Hierarchical linear model · Smart energy · Service platform

## 1 Introduction

The deep integration of Internet technologies such as the Internet of Things, big data and cloud computing with information technology and digital technology has brought about profound changes in the energy system. The deepening of electric power system reform and diversified market service demand further promote the transformation of energy enterprises into intelligent energy comprehensive service providers. The strategic goal of “three types and two networks” put forward by State Grid Corporation of China, as well as the strategic deployment of ubiquitous power Internet of Things, pose new challenges to power enterprises. The development of smart energy integrated services has become a growth point to improve energy efficiency, expand new business, and become an important development direction to promote competition and cooperation.

Many countries comprehensive service platform for intelligent energy research started earlier, the European focus on a variety of energy cooperation between collaborative optimization of complementary, energy system has more exploration, such as Germany's e-energy project is mainly to carry out the energy system and



communication between information system integration, including smart power, smart grid, smart consumer and smart energy storage four aspects; The typical integrated energy system (IES) development plan in the United States aims to increase the proportion of clean energy supply and utilization, and further improve the reliability and economy of the social energy supply system. Japan focuses on optimizing its energy mix and improving energy efficiency, while promoting the large-scale development of renewable energy. At present, China has also accelerated the research and application of comprehensive energy services. Traditional energy enterprises and market-oriented energy enterprises have set up their operations in transportation, medical care, airports, large industrial parks and communities to carry out research and application of platform construction and business model exploration.

At this stage, the smart energy integrated service platform collects business data and optimizes the platform structure framework, so as to better promote the construction of new energy. In order to ensure the data reliability of the smart energy integrated service platform, it does a good job in external data acquisition and management, and analyzes the platform data sharing service and data reliability mining, as well as the platform data operation reliability evaluation and detection [1]. In order to better build the energy big data sharing and operation service platform and ensure the safe and stable operation of the platform, it is necessary to realize the safe processing of the whole chain data with the support of data standardization management, so as to enhance the agile development ability of the intelligent energy integrated service platform and effectively improve the reliability evaluation effect of the intelligent energy integrated service platform [2].

## **2 Reliability Evaluation of Smart Energy Integrated Service Platform**

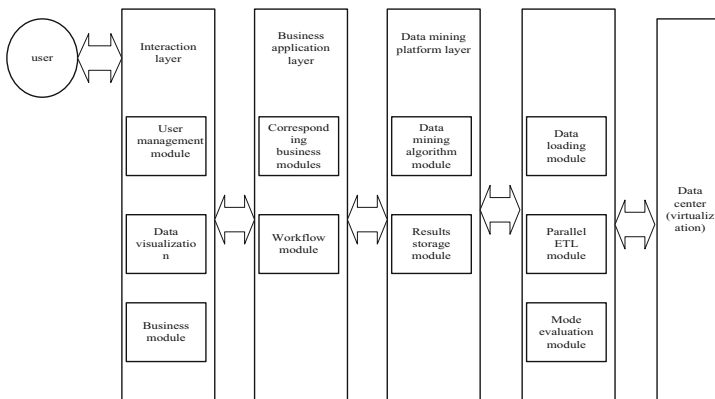
### **2.1 Information Management of Smart Energy Integrated Service Platform**

The reliability evaluation of intelligent energy integrated service platform is the core content of reliability engineering. According to the structural relationship and failure data of each component of intelligent energy integrated service platform, the level of existing platform reliability is quantitatively estimated and evaluated [3]. According to the different evaluation objects, the reliability evaluation methods are divided into: intelligent energy integrated service platform reliability evaluation, platform reliability evaluation and service reliability evaluation, and the evaluation methods, means and tools used by different evaluation objects are also different [4]. Data reliability mining of intelligent energy integrated service platform is the core of energy prediction and energy information discovery based on intelligent energy service. After the preliminary processing steps, the target data set is transformed into the data form suitable for mining. Based on this, the hierarchical linear model is used to mine the energy data to be mined on the platform. The hierarchical linear model is formed by the continuous iteration of the multi-layer linear model. The whole iteration process is a machine learning process. For example, the energy data will continue to iterate until all the

residuals are equal to or close to 0. Based on this, we can effectively form a high accuracy intelligent energy prediction model [5]. Using smart energy services to predict energy data has high efficiency and accuracy, which can well realize the energy demand forecast proposed by users at the initial stage of energy information discovery.

Using hierarchical technology to mine the security data of intelligent energy integrated service platform, in the process of mining, each level becomes very independent, especially easy to expand. Using the asynchronous computing function of intelligent energy integrated service platform, the distributed storage of files is realized.

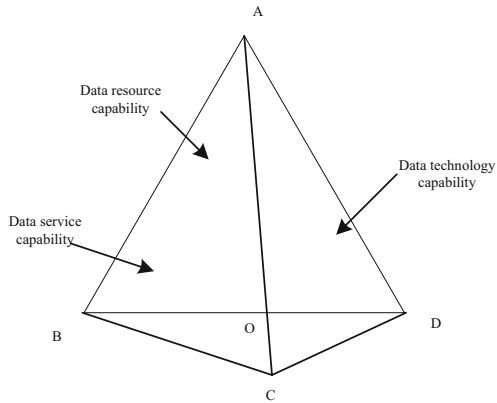
The following is a very detailed introduction of cloud service platform for data reliability mining based on intelligent energy integrated service platform [6]. In the distributed computing platform layer, the intelligent energy integrated service platform is used to complete the data storage function and calculation and analysis function of the cluster. The intelligent energy integrated service platform controls the distributed platform and provides the distributed file platform and parallel operation mode. At the same time, we need to submit the task to the server node. The data reliability mining platform layer is the top priority of the whole architecture, and its general task is to realize the parallel computing of the algorithm. In the process of executing various tasks, the tasks are assigned to the distributed computing layer of the intelligent energy integrated service platform for mining calculation [7]. And the results are fed back to the business application layer. It provides the business application layer with the data of each module needed in each stage of mining business process, including the data collection process, data reliability mining process, pattern evaluation process, data preprocessing process, result display process and other functional modules [8]. Based on this, the overall structure of big data external cloud services is optimized, and the specific structure of smart energy integrated service platform security data mining model is shown in the figure below.



**Fig. 1.** Security data mining model framework of smart energy integrated service platform

As can be seen from Fig. 1, during the operation of the smart energy integrated service platform, it is affected by the complex environment, weak risk tolerance and other factors by the way of alliance sharing to obtain core data, so the possibility of better obtaining core competence is reduced. However, through the flow, transfer,

learning and other processes of energy information among alliances, the possibility of better obtaining core competence is reduced, quickly realize the complementary of the collected data to ensure the information security. According to the concept definition and elements of enterprise core competence proposed by panhand and hammer, the mining data is screened and classified, and the platform reliability evaluation architecture is optimized according to data resource capability, data technology capability and data service capability, as shown in Fig. 2.



**Fig. 2.** Platform reliability evaluation architecture

In the actual operation process, the smart energy integrated service platform has experienced the decentralized development stage *A*, bilateral cooperation stage *B*, multilateral cooperation stage *C*, alliance network stage *D* and cross-development stage *O*, and each stage has different characteristics at the base level, organizational level, management level and information level. It is reflected in the network core, network node, organization mode, cooperation mode, management mechanism, information communication and so on. Therefore, the platform data processing capacity is relatively large. Therefore, the hierarchical linear model is needed to be combined for data processing. The specific platform data processing data is shown in Table 1 below:

**Table 1.** Characteristic information management of smart energy service platform

Stage category	Bilateral cooperation stage	Multilateral cooperation stage	Decentralized development stage	Alliance network stage
Information layer	Simple docking	Interconnection	Nothing	Building information platform
Organizational layer	Bilateral cooperation, limited number of members	Multilateral cooperation with a large number of members	Single member	Multilateral cooperation with a large number of members
Management layer	Limited scope of cooperation and standardized management	Risk sharing control, relationship management and benefit distribution mechanism	Nothing	More attention should be paid to the synergy effect, stability and strategic execution effect of the alliance
Foundation layer	Simple dual core network	Multi core network with various organizational forms	No network or simple single core network	Multi level network structure

The starting point of energy prediction information discovery process is the user, and the user is also the end point of energy information discovery [9]. The whole information mining and management process of smart energy platform needs to be carried out around the users, so the energy information discovery starts from the users. Through the analysis of user needs, the task and target of energy information discovery as well as the type of source data are clarified, so as to simplify the complexity of energy information discovery source data collection and help to locate the data source more effectively and accurately.

### 2.2 Intelligent Energy Integrated Service Reliability Data Optimization Algorithm

The energy prediction and energy information discovery process in smart energy service platform is different from the traditional information discovery process. It is a prediction and discovery process for energy data based on new network environment and high-density electronic data [10]. On the basis of following the general rules of information discovery based on smart energy service platform, the flexible and robust advantages of hierarchical linear model are used to effectively consider the problem of how to select huge data sources on the Internet, and the conversion of initial data source format is added in the process of energy information discovery, which improves the compatibility of energy information discovery process [11]; The predictive energy information contained in the data source is effectively found through the gradient elevator, and the feedback mechanism is added in the platform information discovery process, forming a set of closed-loop and feasible operation for the predictive energy information. The specific process is shown in Fig. 3.

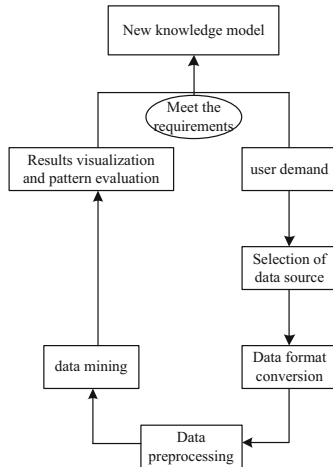


Fig. 3. Steps of reliability information stability prediction of service platform

In the process of evaluating the reliability of smart energy integrated services, we need an additive model composed of multiple weak classifiers in the platform. By accumulating several basic classifiers, the optimized objective function and the exponential reliability loss function related to the class spacing distribution are trained in the gradient direction [12]. The strong prediction parameter  $h(x)$  is constructed by the ensemble prediction model. In the  $m$  steps of gradient promotion, if there is some weak interference  $F_m$  in the smart energy integrated service platform, which will affect the reliability of the platform, the platform information prediction model is calculated by adding a predictor. The details are as follows:

$$F_{m+1}(x) = F_m(x) + h(x) - m \quad (1)$$

The way to find the platform information is as follows:

$$R = SF_m(x) + h(x) = F_{m+1}(x) - y \quad (2)$$

In the above algorithm,  $y$  is the eigenvector and  $S$  is the interference value. The results are as follows:

$$W = Sy - RF_m(x) \quad (3)$$

The approximate function of the reliability of the selected data is further calculated [13]. The specific algorithm is as follows:

$$F_{ij}(x) = \arg WR \min_F E_{x,y}[L(y, F(x))] = \sum_{i=1}^n \gamma_i h_i(x) + WRconst \quad (4)$$

In the above formula,  $W$  represents the reliability loss coefficient, and  $R$  represents the calculation error.  $\min_F$  represents the minimum value of weak interference,  $E_{x,y}$  represents the sample data processing function,  $L$  represents the average loss of sample data,  $\gamma_i$  represents the feature vector of sample data, and  $h_i(x)$  represents the reliability evaluation function of sample data [14].

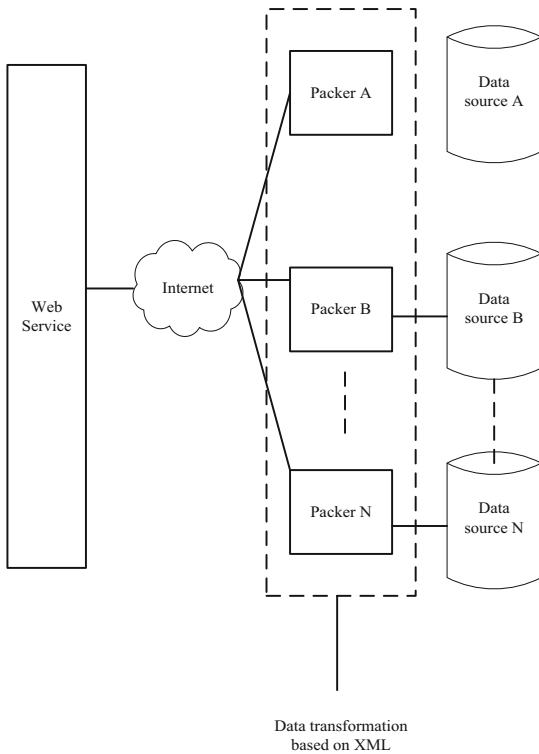
Furthermore, the average loss limit  $L(y_i, \gamma)$  of the information training set samples collected in the smart energy service platform is standardized, and the greedy strategy expansion model is further used to expand the loss function:

$$F_0(x) = F_{ij}(x) \operatorname{argmin}_{\gamma} \sum_{i=1}^n L(y_i, \gamma) \quad (5)$$

$$F_m(x) = F_{m-1}(x) + F_0(x) \operatorname{argmin}_{h \in H} \sum_{i=1}^n L(y_i, F_{m-1}(x) + h(x_i)) \quad (6)$$

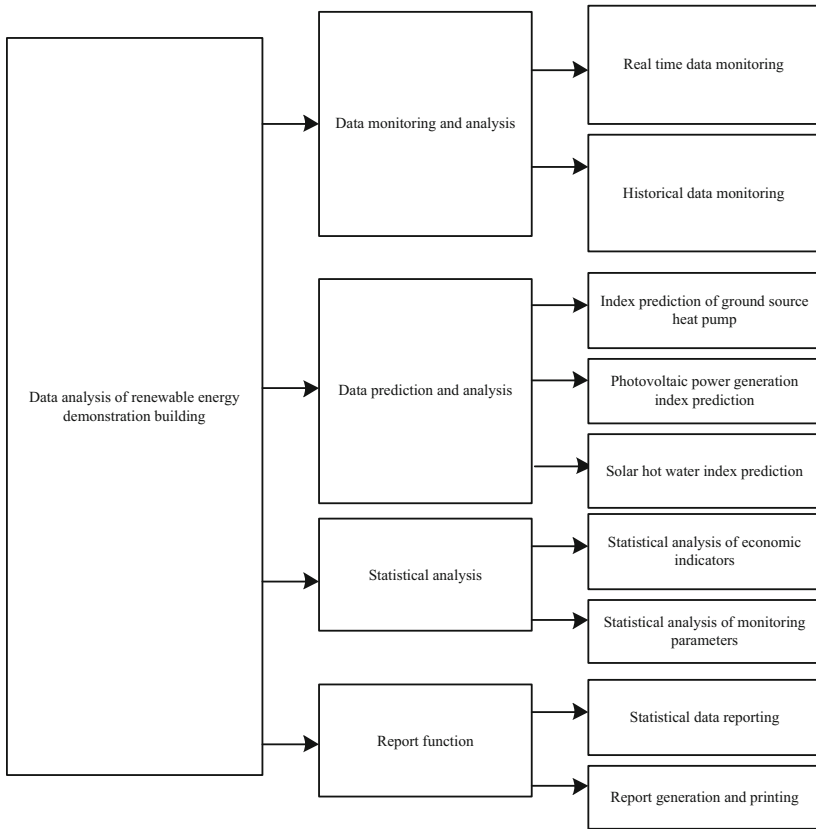
Intelligent energy integrated service platform is a complex data analysis platform which integrates multiple application sub platforms. Its main task is to build a data storage, processing and network publishing platform of web service monitoring center

based on XML, receive all the monitoring data of existing experiments, and store them in the database [15]. The heterogeneous data conversion platform is developed to convert the original heterogeneous data of different regions and different resource climate regions into XML format. By comparing and analyzing the original data of each monitoring point, the smart energy evaluation model of the contribution rate of smart energy to energy consumption is established to comprehensively evaluate the use of equipment and resources in the climate zone, and provide the function of network publishing to provide the latest data and historical data browsing, data import and export, data collection and management different data, various forms of graphics and tables display and report functions [16]. Based on this, the topological structure of platform reliability information evaluation is optimized, and the specific structure is shown in Fig. 4 below.



**Fig. 4.** Topology

On this basis, the function of the intelligent energy integrated service platform is optimized, and the specific function structure is shown in the following Fig. 5.



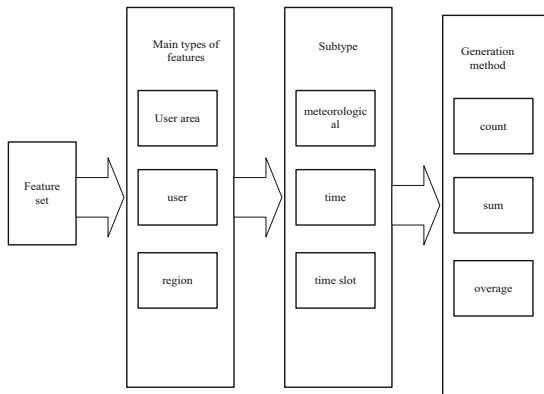
**Fig. 5.** Function structure optimization of smart energy integrated service platform

In order to realize the above functions, considering the use of intelligent energy integrated service platform and project progress, and considering the database and development tools, we decided to use SQL Server database and C# language as the development environment. Before building the energy reliability prediction energy information discovery model based on intelligent energy service, the model construction should follow the construction principle of the model to ensure the smooth progress of the process of energy information discovery [17]. The construction principles of the model mainly include the basic principles and design principles.

### 2.3 Optimization of Reliability Evaluation Algorithm for Smart Energy Service Platform

In the process of energy prediction, feature engineering is needed. The so-called feature engineering is the process of using the relevant energy information in the data field to make the hierarchical linear model achieve the best performance. Feature engineering is an art, and its selection is the key to the success of hierarchical linear model prediction.

Therefore, before data reliability mining, we need to design a  $K$  dimensional feature vector for the original data set. Selecting different feature vectors will usually represent different information sets, which will affect the performance of the whole smart energy service on energy prediction. If the information represented by the smart energy service feature engineering can accurately meet the requirements of the prediction dependent variables, the reliability prediction performance of the model will be improved, otherwise it will be reduced. Therefore, in the intelligent energy service feature engineering of energy prediction and energy information discovery, we need to pay attention to the following principles: whether the feature is emitted or not. For example, in the prediction of power consumption, the variance of a feature is close to 0, that is, there is basically no difference in the original data on this feature, so this feature is not emitted. Whether the features are related or not, that is, the correlation between the features and the prediction target, the features with high correlation should be selected first. To sum up, feature engineering has the advantages of improving prediction accuracy, avoiding over fitting and reducing irrelevant features. According to the above steps and principles of feature engineering, a feature framework is designed, as shown in the figure. Generally, energy prediction is closely related to energy users, regions and other factors, so the characteristics are constructed according to the interaction characteristics of users, regions and user regions (Fig. 6).



**Fig. 6.** Platform reliability information processing structure

In the process of smart energy service modeling, historical data is used to build models to analyze and test the lag effect of time series data. The specific modeling process is as follows: initialize the estimated values of all samples of energy data in category  $K$ ,  $f_{k0}(x)$  represents a vector of  $k$ , specifically represents the estimated values of energy sample data  $x$  in category  $k$ , where  $k$  is also the  $k$  feature of energy processing engineering, then:

$$\{F_m(x) - f_{k_0}(x) \geq 0, k = 1, 2, 3, \dots, K\} \tag{7}$$



By making logistic regression transformation for each estimated sample value,  $f_{k0}(x)$  is converted into the probability value between 0 – 1, and repeated calculation  $K$  times, the reliability probability sequence is obtained:

$$P_k(x) = \exp(f_k(x)) / \sum_{l=1}^K \exp(F_l(x)), \quad k = 1, K \tag{8}$$

In the above formula,  $f_k(x)$  represents the current classification estimate and  $F_l(x)$  represents the logistic regression transformation function.

Traverse the probability of each category in the energy data sample, find the gradient of each sample on category  $K$ , and establish the cost function, which is expressed as:

$$L(\{y_k, f_k(x)\}_1^k) = - \sum_{k=1}^K y_k \log p_k(x) \tag{9}$$

In the above formula,  $y_k$  represents the sample gradient.

Through derivation, the hierarchical linear model is used to learn the energy data, and the gradient form, namely residual form, is obtained. If the energy data belongs to class  $i$ , then residual = true probability - estimated probability. Based on this,  $R$  square, mean absolute error and mean absolute percentage error are introduced as the error measurement index of energy consumption prediction

$$R_2 = \frac{\sum (y_{pre} - y_{avg})^2}{\sum L(\{y_k, f_k(x)\}_1^k) - (y_{ori} - y_{avg})^2} \tag{10}$$

$$MAE = \frac{1}{R_2 n_{sam}} \sum_{i=0}^{n_{sam}-1} |y_{pre,i} - y_{avg,i}| \tag{11}$$

$$MAPE = R_2 \sum_{i=1}^{n_{sam}-1} \left| \frac{y_{pre,i} - y_{avg,i}}{y_{avg,i}} \right| \tag{12}$$

where  $y_{pre,i}$  and  $y_{pre}$  represent the predicted value of energy consumption,  $y_{avg}$  and  $y_{avg,i}$  represent the mean value of sample data, and  $n$  represents the number of samples. The value of  $R_2$  is between 0 – 1, and the larger the value is, the better. However, MAE, MAPE will produce different values according to the data of training set, and the smaller the value is, the better. The probability of no failure operation in the specified operation environment and within the specified time. The reliability of the platform is represented by the function  $R(t)$ , the time of the failure of the platform is represented

by the random variable  $t$ , the probability density function of the random variable  $t$  is represented by the function  $F(t)$ , and the probability of the failure of the program from the initial time to  $t_1$  is represented by  $\Pr 0 < t < y$ :

$$F(t) = \int_0^t \text{MAE} - f(x)dx \tag{13}$$

$$f(t) = \text{MAPE} - \frac{dF(t)}{dt} \tag{14}$$

From 0 to  $t_1$ , the probability of program failure is expressed by  $Pf(t_1)$ :

$$Pf(t_1) = P - f(t)(0 \leq t \leq t_1) = F(t_1) - F(0) - f(t) = F(t_1) - f(t) \tag{15}$$

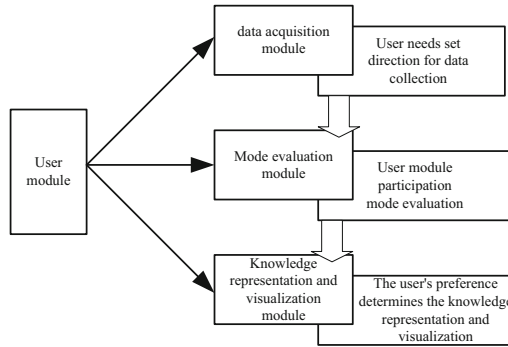
If the probability of successful operation of the program in interval  $[0, t]$  is  $Ps(t)$  and the probability of failure is  $Pf(t)$ , then:

$$\sum F(t_1)Ps(t) + Pf(t) = e \tag{16}$$

In the above formula,  $e$  represents the overall error. It can be deduced from the above formula that:

$$R(t) = 1 - Pf(t)e = 1 - F(t)e = 1 - \int_0^t f(x)edx \tag{17}$$

The definition shows that the platform reliability is defined as a probabilistic measure, and the probability of no failure operation should be used to describe the reliability. The platform reliability is a function of time. When the residual error value is fixed, the longer a platform runs, the greater the probability of failure and the smaller the reliability of the platform. Because reliability is affected by many factors, it is difficult to evaluate reliability with one parameter. For different applications, there may be different parameters. For the evaluation of platform reliability, it can be evaluated from the aspects of management reliability and technical reliability. Management reliability is mainly for management personnel, and less considered in the evaluation of platform reliability. The evaluation indexes of platform reliability are mainly elaborated from the perspective of technical reliability. The indexes commonly used to characterize platform reliability are reliability, failure efficiency, failure strength, reliability, etc. Based on this, the platform reliability information management method is further optimized, as shown in Fig. 7.



**Fig. 7.** Platform reliability information management method

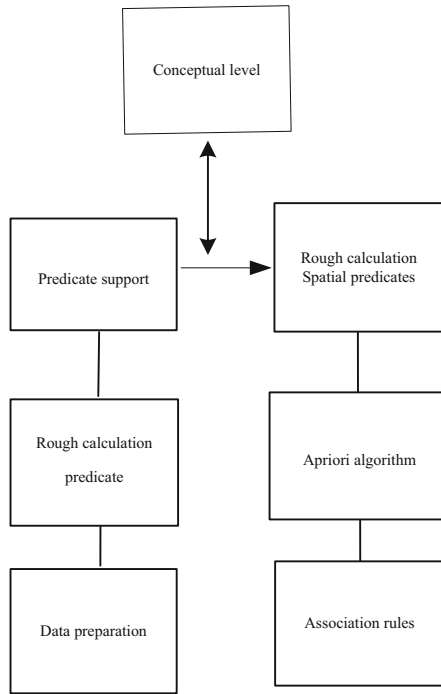
Data preprocessing based on the above methods is mainly the preliminary processing of the source data, which mainly includes two steps: data cleaning and integration and data selection and transformation. Among them, data cleaning is mainly to fill missing values in data sources, correct inconsistent data and remove data noise. The data integration is when the data comes from multiple data sources, it needs to integrate the data, that is, simply integrate and overlay the data, integrate the data from different data sources into a data set, and improve the consistency and accuracy of the source data. The data preprocessing module is the data processing stage of the energy prediction and energy information discovery model based on smart energy services. It primarily processes and analyzes the collected data by means of data conversion, data cleaning, data aggregation, data integration and data specification, so as to form an orderly data warehouse and make full preparation for data reliability mining. This module is mainly aimed at the noise, missing and other problems of the target data set, using the corresponding technical means to process the target data set, so as to form a suitable data warehouse. Therefore, data conversion, data cleaning, data aggregation, data integration and data specification technologies are used in this module, as shown in Table 2.

**Table 2.** Data reliability preprocessing

Data reduction	Simplified representation of data sets
Data aggregation	Simple classification of messy data
Data conversion	The format of the source data set is converted uniformly
Data integration	Merge multiple data
Data cleaning	Filling missing values, identifying discrete points and so on

Platform data selection is to select the appropriate sample from the source data set, and the sample analysis can produce almost the same analysis results as the source data set. Data transformation is realized through discretization, standardization and other

technologies to ensure that data can be mined from multiple abstract levels, which is conducive to improving the quality of data reliability mining. Data reliability mining is the core of this model. The core algorithm of this model is hierarchical linear model, and the content of energy information discovery in this model is time series data, that is, predictive data. Through the analysis of the status and trend of historical data, we find the change trend of future data. Comparing the content of the energy information discovery with the actual data in the first mock exam, it is found that the deficiency of the model is feedback to the former module, and the energy information discovery result is optimized. Based on this, the reliability correction process of platform reliability data is optimized, and the specific process is shown in Fig. 8.



**Fig. 8.** Reliability correction process of platform reliability data

After data query and analysis, it is necessary to collect the target related objects and reference sets into the database. The expected confidence of  $X \Rightarrow Y$  is the percentage of all things contained in association rule  $X \Rightarrow Y$  in database; The confidence is compared with the expected confidence

$$D = \frac{R(t) + X \Rightarrow Y}{\text{expected value: } X \Rightarrow Y} \tag{18}$$

Interestingness can measure the X, Y relevance of all things. Predicate calculation is carried out at the rough level. The target is set as the minimum rectangle, the extraction distance falls within the predetermined threshold as the object, the predicate of object relationship is stored in the database, and the attribute value is set as a single value or a group of values. Different predicates have different support

$$\xi = D \times (X - Y) \quad (19)$$

The threshold with lower support is excluded, and then the common database is formed. In order to perform accurate spatial calculation in common databases, MBR technology is used to check the relationship between predicates, exclude the predicate relationship that does not conform to the actual situation, and then form the topological data table, so as to calculate the support of predicates, exclude the items with less support, and then form the optimal database. Through the model evaluation, we can make appropriate trade-offs on the results of energy information discovery, and can feed back the poor results of energy information discovery, so as to conduct a new round and more detailed energy information discovery on this part of energy information. In the hierarchical variable weight cloud service reliability evaluation model, the reliability of service reflects the reliability of users getting the best service resources. The greater the reliability of service, the better the quality of cloud service; The mathematical expectation of service reliability reflects the average value of service reliability provided by service resources when users get service; The mean square error of service reliability reflects the deviation degree between the reliability of service resources and the mean value of service reliability. The greater the fluctuation is, the worse the effect of reliability is. Therefore, the reliability quality of cloud service can be evaluated by one or several aspects of reliability, mathematical expectation of reliability and mean square error of reliability. According to the platform analysis idea of holism and reductionism, the reliability evaluation model based on hierarchical variable weight realizes that the weight of reliability factors changes with the change of cloud service status. According to different actual needs, it can use incentive type and punishment type state variable weight to build a reasonable reliability weight adjustment mechanism, which provides an efficient and effective way for cloud service reliability evaluation flexible technical means, with strong operability and flexibility.

### 3 Analysis of Experimental Results

In order to verify the feasibility and accuracy of the evaluation method of intelligent energy integrated service platform based on hierarchical linear model, the experiment is divided into 5 user demand analysis, data acquisition, data preprocessing, data reliability mining, data visualization and pattern evaluation. The environment of this experiment is Mac-OS10.13 platform, and the environment is PyCh development environment experiment. It calls the Scrapy crawler of python2.7 to collect data, and pandas database is used for data processing, scikit learn library for linear regression fitting, numpy database for scientific calculation, Matplotlib database for visualization and result export. The hardware environment is 2.7 GHz IntelCorei5CPU and 8 GB

RAM. With the change of parameters, the error amplitude of the model decreases. Based on this, the trend of platform data reliability is investigated. The specific results are shown in Fig. 9 below:

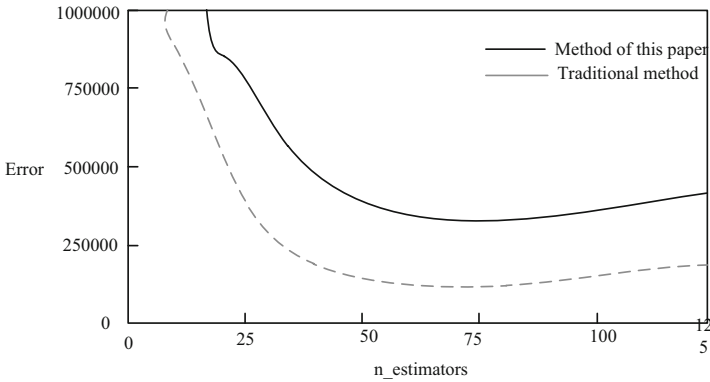


Fig. 9. Trend of platform data reliability

Generally, it is necessary to understand the influence of the constructed features on the model prediction. Therefore, six features, such as summer, minute, Dow, day, year, etc., are selected to draw the feature importance diagram. It can be seen that the summer feature has the least influence on the model, and the influence of minute and doy on the model is relatively large (Fig. 10).

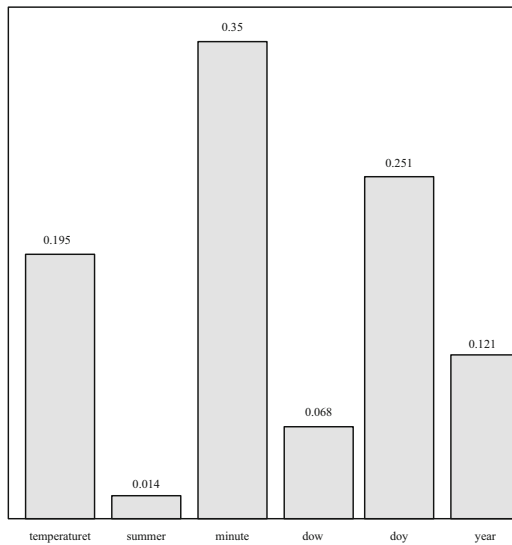
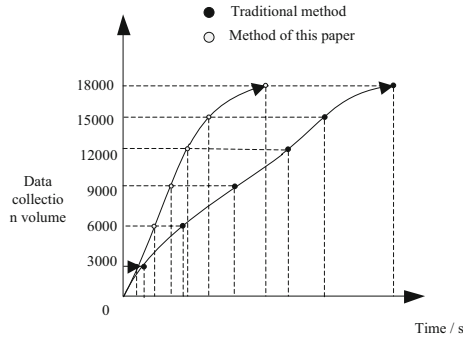


Fig. 10. Influencing factors and influencing degree of platform stability

Through the above steps, this experiment draws the final visualization results, and compares the reliability of data mining speed. The specific detection results are shown in Fig. 11.



**Fig. 11.** Comparison of platform reliability evaluation efficiency results

It can be seen from the figure that the reliability and efficiency of the smart energy service platform proposed in this paper is obviously higher than that of the traditional platform, which can better guarantee the rational use of smart energy and avoid the waste of resources.

Taking the reliability evaluation accuracy and evaluation time of the comprehensive smart energy service platform as experimental indexes, the actual application effects of the traditional method and the method presented in this paper are compared. The specific experimental results are shown in Table 3.

**Table 3.** Comparison of practical application effects

Number of experiment	Evaluation accuracy/%		Evaluation time/s	
	Traditional method	Method of this paper	Traditional method	Method of this paper
1	85.6	98.6	1.9	0.5
2	81.3	99.3	1.8	0.2
3	82.4	98.9	1.4	0.4
4	86.3	97.8	1.6	0.2
5	84.1	98.5	1.7	0.3
6	82.5	98.6	1.5	0.4
7	81.7	98.4	1.2	0.5
8	85.6	97.4	1.8	0.3
9	84.7	97.6	1.4	0.2
10	82.6	98.3	1.7	0.3
Average	83.7	98.3	1.6	0.3

It can be seen from the analysis of Table 3 that the average reliability evaluation accuracy of the comprehensive intelligent energy service platform of the traditional method is 83.7%, and the average reliability evaluation accuracy of the comprehensive intelligent energy service platform of this paper is 98.3%, which is much higher than that of the traditional method. The reliability assessment time of the integrated intelligent energy service platform of the traditional method is 1.6 s, while that of the proposed method is 0.3 s, much lower than that of the traditional method. The above data prove that the proposed method has better comprehensive performance and better practical application effect.

## 4 Concluding Remarks

In recent years, the application requirements of the comprehensive intelligent energy service platform have been increasing. In order to better realize the whole process management of smart energy service projects, realize the objectives of data visualization, online monitoring and early warning, and effective response, and with the help of mobile data collection, big data analysis and index management, etc. The reliability evaluation method of intelligent energy integrated service platform based on layered linear model is proposed in order to provide users with new service experience. Finally, through energy efficiency improvement, energy saving transformation, creating stable revenue, expanding profit source, the ultimate goal of bringing high efficiency, clean, reliable, convenient personality and intelligent interaction to customers in the whole society can be achieved.

## References

1. Abiri, R., Borhani, S., Kilmarx, J., et al.: A usability study of low-cost wireless brain-computer interface for cursor control using online linear model. *IEEE Trans. Hum.-Mach. Syst.* 1–11 (2020)
2. Yang, W., Chen, Y., Chen, Y.C., et al.: Intelligent agent-based predict system with cloud computing for enterprise service platform in IoT environment. *IEEE Access* 1 (2021)
3. Zhang, C.: Design and application of fog computing and Internet of Things service platform for smart city - ScienceDirect. *Futur. Gener. Comput. Syst.* **112**(4), 630–640 (2020)
4. Lei, J., Zhou, C., Li, X., et al.: Energy management considering energy storage and demand response for smart energy hub in Internet of Things. *IEEE Access* 1 (2020)
5. Carli, R., Dotoli, M.: Decentralized control for residential energy management of a smart users' microgrid with renewable energy exchange. *IEEE/CAA J. Automatica Sinica* **6**(3), 36–51 (2019)
6. Li, Z., Li, H., Feng, Y.: Research on big data mining based on improved parallel collaborative filtering algorithm. *Clust. Comput.* **22**(5), 1–10 (2019)
7. Arafat, Y., Hoque, S., Xu, S., et al.: Machine learning for mining imbalanced data. *IAENG Int. J. Comput. Sci.* **46**(2PT.264–383), 332–348 (2019)
8. Campo-Vila, J.D., Takilalte, A., Bifet, A., et al.: Binding data mining and expert knowledge for one-day-ahead prediction of hourly global solar radiation. *Expert Syst. Appl.* **8**, 114147 (2020)



9. Nali, J., Martinovi, G., Agar, D.: New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers. *Adv. Eng. Inform.* **45**(11), 101130 (2020)
10. Lin, M., Chen, X., Wang, Z., et al.: Global profiling and identification of bile acids by multi-dimensional data mining to reveal a way of eliminating abnormal bile acids. *Anal. Chim. Acta* **32**(11), 74–82 (2020)
11. Lu, N.: Reliability evaluation model of network resource sharing information consulting service platform. *Environ. Technol.* **37**(2), 45–48+52 (2019)
12. Fei, Q.: SOA complex network system reliability evaluation. *Comput. Technol. Dev.* **29**(11), 113–117 (2019)
13. Jiang, J.L., Guan, T.T.: RESTful service reliability evaluation optimization based on Markov chain. *Mod. Comput.* **23**(8), 17–21 (2018)
14. Xuan, J.J., Xiao, Y., Lu, W.S., et al.: Reliability evaluation method for regional integrated energy system applying complex network theory. *Electr. Power Constr.* **41**(4), 1–9 (2020)
15. Lu, Z.K., Jian, X.H., Zhang, M.H.: Reliability and economic evaluation of multi-terminal flexible DC distribution network. *South. Energy Constr.* **7**(4), 67–74 (2020)
16. Pan, J., Shen, Y.Y., Duan, R.C.: Interval estimation of power supply reliability evaluation for planned power outage projects. *Power Energy* **40**(4), 62–66 (2019)
17. Ni, W., Lv, L., Xiang, Y., et al.: Reliability evaluation of integrated energy system based on Markov process Monte Carlo method. *Power Syst. Technol.* **44**(1), 150–158 (2020)



# Design of Clinical Medical Data Monitoring System Based on Artificial Intelligence and Big Data

Tao Lei<sup>1</sup> and Gui-xiu Xie<sup>2</sup>(✉)

<sup>1</sup> Physical and Electronic Engineering School, South China Normal University, Guangzhou 510006, China

leitao022254@aliyun.com

<sup>2</sup> Guangdong Polytechnic College, Zhaoqing 526100, China  
xx210428@163.com

**Abstract.** With the development of medical information technology, the scale of clinical medical data increases exponentially, which brings great difficulties to the management and monitoring of clinical medical data. The hardware unit of clinical medical data monitoring system consists of hardware frame unit, ARM processor selection unit, RAM memory selection unit and its gateway and remote client device selection unit. Through the design of hardware unit and software module, the operation of clinical medical data monitoring system is realized. Compared with the existing system, the design system has shorter response time and higher monitoring efficiency, which fully proves that the design system has better application performance and provides an effective means for hospital administrators and public health management departments.

**Keywords:** Artificial intelligence · Big data · Clinical medicine · Medical data · Monitoring

## 1 Introduction

Big data has opened a major era of transformation. Its concept in recent years has become more and more recognized, and has applied to all walks of life in scientific research production. It is no longer far away in people's lives. Engineers at Google used years of search-log data to predict the spread of winter flu in the weeks before the H1N1 outbreak, and to predict specific regions and states; Farecast used flight records and airfare data to predict fares for domestic flights with 75% accuracy; and Xoom successfully spotted criminal gangs trying to defraud each transaction by analyzing data from cross-border remittance companies. Big data is changing the world, from technology to business, economics, humanities, education, health care and other aspects. Big data holds great value, the application of Big data makes us face great opportunities which will bring great change to the whole era.

In the medical and health industry, but also because of its industry characteristics, clinical medical data has been more and more mentioned in recent years, become the object of study. But how to make "Big Data of Clinical Medical Treatment" from the

“cloud” to be a real and effective information product or service is still something worth studying. The clinical medical industry has met the challenge of massive data and unstructured data for a long time. Because of the limitations of technology and finance, it has not been applied well. In recent years, many countries are actively promoting the development of medical informationization, as well as artificial intelligence technology, which means many medical institutions have the funds and direction to do big data analysis. With the rise of electronic health records systems, imaging systems, electronic prescription software, medical claims, public health reports, and related applications and mobile medical devices, the medical services industry has emerged as the primary beneficiary of big data with the greatest analytical potential. Huge amounts of data from medicine, pharmacy practices, health care professionals, patients, medical records, and even institutions are waiting for big data analysis. The analysis of these data, which is used as a reference for the treatment of clinical patients, has begun to serve as an entry point for the analysis and application of medical big data, with a growing number of institutions and scholars conducting research [1]. On the other hand, how to use this medical data in the supervision of medical behavior and medical quality from the perspective of public management of clinical health care is seldom discussed.

With the development of hospital management and the improvement of hospital supervision, it is imperative to strengthen hospital supervision by using modern information. But at present, the information construction of domestic hospitals mainly focuses on business process management and charge management, and seldom involves the monitoring of clinical medical data quality. For example, reference [2] rethinks the issue of patient consent in the era of artificial intelligence and big data. However, there is not much mature experience for reference in monitoring the quality of clinical medical data through the information system, including whether the doctor’s diagnosis and treatment is in line with clinical norms, excessive or lack of medical treatment, whether the medication is safe, and whether the diagnosis and treatment of patients with medical insurance is in line with the provisions of medical insurance policy.

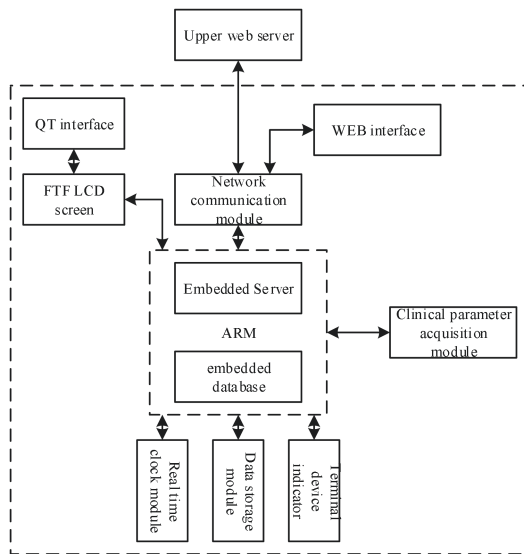
Based on the analysis of medical big data and artificial intelligence technology, this paper collected, calculated and analyzed clinical medical data, and found these irrational and irregular phenomena. According to the above ideas, a set of clinical medical data monitoring system is designed and implemented in medical institutions and public health management departments, which provides effective means for hospital managers and public health management departments.

## **2 Hardware Unit Design of Clinical Medical Data Monitoring System**

The hardware unit is the basis for designing the stable operation of the system. In order to solve the problems existing in the existing system, the hardware unit of the system is designed, including hardware frame building unit, ARM processor selection unit, RAM memory selection unit and its gateway and remote client device selection unit. The specific design process is as follows:

## 2.1 Hardware Framework Building Unit

The system adopts modular design method, adopts ARM system V5 architecture S3C2440A as the main control unit chip, and the hardware design is divided into the functional module design of the terminal main control board and the interface board, mainly including RAM memory circuit, Nand Flash memory circuit, power supply module circuit, ECG rate collection module circuit, body temperature parameter collection module circuit, pulse oxygen saturation collection module, etc., and adds the LED lamp to display the system status [3]. The system functional structure diagram is shown in Fig. 1.



**Fig. 1.** Design system hardware framework diagram

Based on the design system hardware framework mentioned above, the system hardware is selected. Due to the limitation of space, this study only scientifically selects ARM processor, RAM memory, its gateway and remote client devices, and other devices continue to use existing system hardware devices [4].

## 2.2 ARM Processor Selection Unit

The S3C2440A is a microprocessor based on the ARM920T core. It is a high integrated, low power processor introduced by Samsung. It is developed with 0.13 um CMOS standard cell and memory compiler. It adopts a new bus architecture known as the Advanced Microcontroller Bus architecture. The core of the S3C2440A has an advanced RISC machine designed with a 32-bit ARM920T RISC processor. Its low-power, simple, and all-static design is particularly suited to cost-and power-sensitive applications [5].

The chip integrates the following components: 16kB instruction cache, 16kB data cache, external storage controller, LCD-specific DMA LCD controller, 4-way DMA controller for external request pins, 2-way SPI, 3-way URAT, 2-way USB host control/1-way USB interval control, 4-way PWM timer/1-way internal watchdog timer/watchdog timer, 8-way IIC bus, 8-way 10-bit ADC and touchscreen interface.

The control circuit needs to realize the function is similar to the host computer in the PC. The control circuit design includes microprocessor, SDRAM, Flash and various other device interfaces, including LCD interface, touch screen interface, USB interface and DART interface etc. [6].

### 2.3 RAM Memory Selection Unit

The S3C2440A microprocessor itself has only 16-kilobyte instruction Cache and 16-kilobyte data Cache, without internal RAM or ROM. In order to make the main control system work normally, it is necessary to configure external RAM memory in the control circuit.

Although the S3C2440A does not have physical RAM memory, the S3C2440A is configured with an external storage control unit that includes the SDRAM controller and slicing logic. The storage controller of the S3C2440A provides the required memory control signals for access to external storage channels [7].

The S3C2440A's storage controller has the following characteristics:

First, support the mode of large and small terminals (software selection);

Second, there are a total of eight memory banks, and Bank0 can be used as a boot ROM, with the exception of bank0, other banks have programmable access sizes.

Third, the address space: bank0- bank5 fixed 128 MB;

Fourth, 7 fixed start address banks, for each bank programmable loop access;

Fifthly, the starting address of bank7 is adjustable, and the capacity of bank6 and bank7 is programmable.

Sixth, support SDRAM automatic refresh mode and power-down mode.

In a S3C2440A reset memory image, the SROM represents ROM or SRAM -type memory, and the SFR represents a special functional register. OM0, OM1 are the two external pins of the S3C2440A, when OM [0: 1] = 0 the processor can boot from Nand Flash, and when OM [0: 1] = 1 the processor can boot from j at 0x00 \_ 0000 instruction unit.

In theory, the entire 32-bit CPU storage size is 4 GB, which is from address 0x000 \_ 0000 to 0xFFFF \_ FF. For S3C2440A, from 0x0000 \_ 0000 to 0x4000 \_ 0000 is the actual address space available. Space is reserved for 0x6000 \_ 0000 to 0xFFFF \_ FFFF. The first 1 GB of S3C2440A is divided into 8 groups of memory blocks, which correspond to the 8 enabling pins of nGCS [0: 7] chip respectively.

Blocks 0-5 are ROM/SRAM blocks, blocks 6 and 7 are SROM/DROM/SDRAM blocks, and when the system is accessed at the corresponding storage location, the corresponding nGCS [0: 7] pins are enabled. From 0x4800 \_ 0000 to 0x6000 \_ 0000 are special function registers for S3C2440A, that is, registers for other function units on the S3C2440A control chip. Because ARM has fixed registers, but different companies produce ARM -based microprocessors will vary, usually there is a specific block of memory to store these control registers storage location.

Because SDRAM is controlled by a storage controller, not every SDRAM chip is supported by the CPU of the ARM. The sixth and seventh chunks of the S3C2440A support 2 MB, 4 MB, 8 MB, 16 MB, 32 MB, 64 MB, and 128 MB SDRAM. The SDRAM storage mapping space address configuration is shown in Table 1.

**Table 1.** SDRAM storage mapping space address configuration table

Block size	Bus width	Memory architecture
32 MB	×16	$(8M \times 4 \times 2 \text{ banks}) \times 4 \text{ ea}$
	×16	$(4M \times 4 \times 2 \text{ banks}) \times 4 \text{ ea}$
	×32	$(4M \times 5 \times 2 \text{ banks}) \times 4 \text{ ea}$
	×32	$(2M \times 8 \times 4 \text{ banks}) \times 4 \text{ ea}$
	×16	$(4M \times 8 \times 4 \text{ banks}) \times 2 \text{ ea}$
	×32	$(2M \times 16 \times 4 \text{ banks}) \times 1 \text{ ea}$
	×8	$(8M \times 8 \times 4 \text{ banks}) \times 1 \text{ ea}$
	×16	$(4M \times 16 \times 4 \text{ banks}) \times 4 \text{ ea}$
64 MB	×32	$(4M \times 8 \times 4 \text{ banks}) \times 4 \text{ ea}$
	×16	$(8M \times 8 \times 4 \text{ banks}) \times 2 \text{ ea}$
	×32	$(4M \times 16 \times 4 \text{ banks}) \times 2 \text{ ea}$
	×8	$(16M \times 8 \times 4 \text{ banks}) \times 1 \text{ ea}$

As shown in Table 1, block size refers to the size of the SDRAM as viewed from the ARM CPU, depending on the maximum access range of the CPU. The maximum 128 MB SDRAM for S3C2440A was described earlier, but the address configurations for SDRAM with 32 MB and 64 MB block sizes are listed here due to space limitations. Bus width is the unit of one-time access by the CPU. Memory architecture refers to the specification of the SDRAM provided by the main control chip memory controller. The architecture of SDRAM includes BA pins, and the target chip uses different address lines for block address selection.

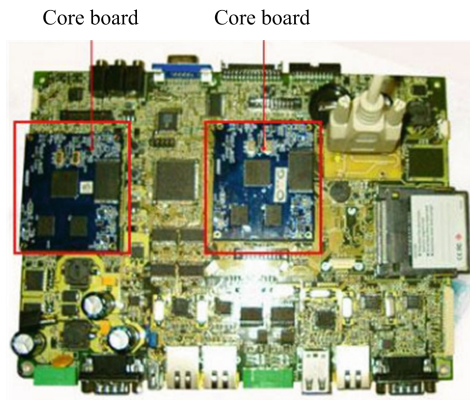
ARM -based microprocessor memory can be divided into two types according to the different access methods. One is asynchronous memory, such as ROM, EPROM, etc. The biggest feature of this type of memory is that the number of address buses can only determine the number of units of memory, suitable for small-capacity system design, which is usually used to place the system's bootstrap program. The other is synchronous memory, such as SDRAM. This type of memory must have a set of synchronized clocks. It runs faster than asynchrony, and its address bus does not correspond to the size of the memory. The SDRAM address bus can represent both a row address bus and a column address bus, and the row and column address buses can be grouped differently depending on the SDRAM [8]. There are also so-called block address choices, and the number of blocks varies depending on the SDRAM, usually divided into 2 or 4 blocks.

In order to improve the system running speed and facilitate future system upgrades, 64MB SDRAM is pre-configured in the control circuit design. As you can see in Table 1, there are two options for configuring a 64 MB SDRAM with a 32-bit bus

width. One option is to select an SDRAM chip with a base unit of 128 Megabytes and a memory architecture of 4 M84banks and 4 EA. The second option is to select an SDRAM chip with a base unit of 256 Mb and a memory architecture of 2 EA (4 M164banks). Since both solutions require four banks, two address lines, A24 and A25, are required to implement block selection.

## 2.4 Gateway and Remote Client Device Selection Unit

The hardware platform used in the gateway is OK6410 development board, which is shown in Fig. 2.



**Fig. 2.** OK6410 development board physical schematic

The resources on the development board are as to

Processor: Samsung S3C6410, ARM11 core;

Main frequency: 553 MHz/667 MHz;

Memory: 256 MB;

NAND Flash: 1 Gbytes;

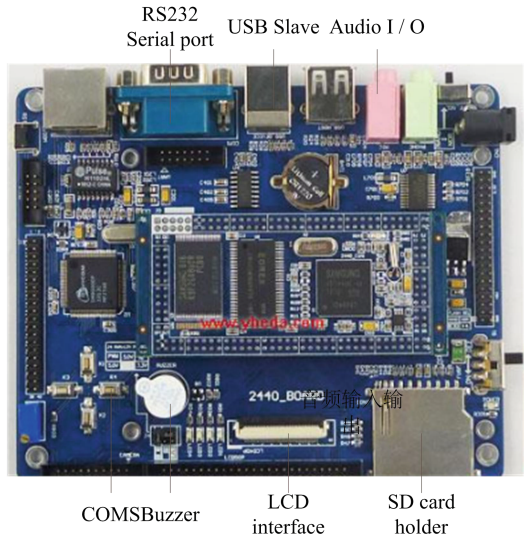
Network Card: 1 100M network port, using DM9000AE, with the connection;

Serial port: 4 serial ports, including a RS232 serial port and 3 T;

USB: A USB HOST and a USB Slave interface;

LCD: 4.3 in. TFT LCD screen.

The hardware platform adopted by the remote client is TQ2440 development board, the physical development board is shown in Fig. 3.



**Fig. 3.** TQ2440 development board physical schematic

The resources on the development board are as follows:

Processor: Samsung S3C2440, ARM9 core;

Main frequency: 400M, can double frequency to 533M;

Memory: 64M;

NAND Flash: 256 M;

Network Card: 1 100M network port, using DM9000AE, with the connection and transmission indicator l;

Serial port: an RS232 serial port and 3 expanded interfaces;

USB: A USB HOST and a USB Slave interface;

LCD: 4.3 inch TFT LCD screen.

The above process completes the design and selection of system hardware, but still can not achieve the monitoring of clinical medical data, so based on this, design system software module.

### **3 Software Module Design of Clinical Medical Data Monitoring System**

The software module of the design system includes data acquisition and transmission module, data mining module and data stability control module. The software module design process is as follows:



### 3.1 Data Acquisition and Transmission Module

The clinical medical data mainly include temperature, ECG rate, pulse oxygen saturation, etc., and the above data collection and transmission schemes are selected as follows:

(1) Temperature measurement scheme

There are two methods to collect body temperature data: traditional contact method and popular electronic instrument method.

Mercury thermometer: This is the most common clinical contact method of measuring body temperature. A glass tube is filled with mercury in its storage sac. Using the principle of mercury expansion by heat, the body temperature is displayed. This method has the advantages of convenience and practicality. But because this kind of thermometer is easy to volatilize after breaking, pollutes the environment, this kind of measurement method has been eliminated in some areas.

Electronic thermometers: Common electronic thermometers are thermistor and thermocouple thermometers [9, 10]. The principle of thermistor thermometer is that the value of thermistor in thermistor varies with heat expansion and cold contraction, and the principle of thermocouple thermometer is that the current varies with temperature in the circuit composed of two different kinds of metals. Because of the rapid development of science and technology, the cost of manufacturing thermometers is gradually declining, electronic thermometer measurement has been quite mature. In the existing technology, electronic thermometer can not only achieve discontinuous measurement, but also can achieve continuous measurement, which provides help for continuous monitoring of body temperature.

At present, there are two methods of human body temperature measurement: contact method and non-contact method. Infrared thermometry is a common non-contact method of temperature measurement based on the principle of thermal radiation. The advantage is that the measurement technology is mature and convenient, and it is a development direction of measurement technology.

The MLX90615 is a contactless electronic chip. Using infrared measurement. The measured data have the advantages of high accuracy and stability. So this system uses non-contact 1VB, X90615 chip for temperature measurement.

(2) ECG rate measurement programme

The commonly used ECG acquisition modules are very complex, including input conditioning circuit, preamplifier circuit, shielding driving circuit, high-pass filter circuit, low-pass filter circuit, etc. In the acquisition process, the ECG signal collected by the input electrode is amplified through the amplification circuit, then the low-frequency signal contained in the ECG signal is filtered through the high-pass filter circuit, and the power frequency signal and the high-frequency signal are interfered with and filtered. Finally, the signal is amplified and converted into a digital signal that is easy to identify through an analog-to-digital converter, and the data is provided to the processor for processing.

Nowadays, more and more electronic technology is used in medical treatment. Many companies have introduced small integrated circuit chips, in which filter circuits, amplifiers and AD converters are concentrated. There is a digital signal processing module inside the chip for digital filtering algorithm. So choose a small BMD 101 ECG acquisition chip, not only can reduce the layout of the circuit, but also can make the design more concise, form a modular design.

(3) Pulse blood oxygen saturation measurement scheme

Common electrochemical methods and optical methods are widely used in the measurement of blood oxygen concentration. Traditional electrochemical methods first need to take blood from the human body, and then electrochemical analysis by an analyzer to get the partial pressure of arterial oxygen, so as to calculate the oxygen saturation of arterial blood. Compared with electrochemical methods, optical methods not only do not need to collect blood, will not cause injury to the human body, but also can be real-time continuous monitoring. The main principle of the photochemical method is to calculate the oxygen saturation of blood by using the different absorption of light by blood at different wavelengths. Although the accuracy of the photochemical method is lower than the electrochemical method, it is simple to use. Its clinical effects have been widely recognized. YS2000A module using photochemical measurement method, high measurement accuracy, good real-time. Therefore, this monitoring system uses YS2000A module with digital blood oxygen probe to collect pulse and blood oxygen saturation.

Human physiological parameters include ECG, heart rate, pulse, oxygen saturation, body temperature and so on. According to the design characteristics of telemedicine data acquisition system, this paper adopts the following design schemes. The ECG module uses BMD101 chip, the oxygen module uses YS2000A module designed by Shanghai Berry Electronics Company and digital oxygen probe to collect pulse and oxygen saturation information, and the body temperature measurement uses MLX9061 chip, so the body temperature measurement can be completed without building other circuits. This modular design facilitates system maintenance, but also reduces the cost of research and development.

The existing data transmission methods are mainly as follows:

**WiFi technology:** WiFi technology is based on the IEEE 802.11 protocol IEEE 802.11 protocol, the Chinese name for wireless fidelity technology. WiFi technology has been rapidly applied due to its unique technical advantages since it came out. It has the advantages of long transmission distance, high transmission speed and no need for wiring. At the same time, WiFi technology allows many users to access, easy to expand the system, has become a rapid development direction of the infinite local area network.

**Bluetooth technology:** Bluetooth technology is widely used in people's daily life, such as Bluetooth headset, Bluetooth speaker and mobile phone data transmission. The Bluetooth Technology Alliance is mainly responsible for the management and certification of Bluetooth projects, the maintenance of Bluetooth certified trademark rights and the effective supervision of Bluetooth protocol development. To enter the consumer market as a Bluetooth device, the electronic device manufactured by the manufacturer must comply with the Bluetooth device specification standards introduced by

SIG. Its characteristics are short transmission distance, slow data transmission rate, suitable for short distance and small capacity transmission. Because of its common frequency band, it is easy to be disturbed by the outside.

Ethernet transmission: The most common standard of communication protocol used between communication devices. Ethernet networks use CSMA/CD technology, and Ethernet transmits packets between interconnected devices at a rate of 10 to 100 Mbps. Ethernet is similar to the IEEE 802.3 standard. It has the characteristics of fast data transmission speed, good confidentiality, can use a variety of transmission media. Suitable for large capacity transmission without distortion.

According to the special requirements of clinical medical data monitoring system, the system chooses Ethernet which has better confidentiality to transmit medical data.

### 3.2 Data Mining Module

Data mining is the application of big data analysis method, which can mine clinical medical data deeply and lay a solid foundation for its monitoring.

Data mining is an interdisciplinary field, which combines database system, statistics, machine learning and other technologies into one. Data mining can be described from the business point of view as: according to the established business objectives of enterprises, exploring and analyzing a large number of enterprise data, revealing hidden, unknown or verifying the known regularity, and further modeling its data processing methods. In the field of medical big data, many data mining techniques have been successfully applied to clinical medicine and scientific research. Based on the basic concepts of data mining, this chapter focuses on some algorithms of data mining, including the most common association rules algorithm and clustering classification algorithm, and applies them to the clinical data of patients, and tries to discover some unknown rules in the clinical data of patients by using the algorithm rules, so as to provide a scientific basis for medical prediction and optimization of medical plan.

The technologies and methods used in data mining mainly come from interdisciplinary and related technical fields, which can be summarized as the following seven points:

- 1) Statistical methods: certain statistical processes are involved in data mining, such as mathematical sampling and modeling, hypothetical judgment, etc. Nowadays, under the background of big data, distributed storage and computing capacity are becoming stronger and stronger, and the sampling calculation method can even be directly evolved into total calculation.
- 2) Decision tree: Decision tree is mainly used for data classification and prediction, and has many implementation algorithms, including ID3 learning algorithm, SLIQ algorithm, etc. In the clinical medical forecast project of a hospital, the decision tree classification method of data mining is applied to the test report data of patients, and the decision tree graph of the test report data of patients is obtained.
- 3) Artificial neural network: This technology mainly imitates biological neural network, and improves the learning ability of the machine through data training, so as to achieve the effect of autonomous prediction. Often used for clustering, feature mining and other operations.

The above mentioned technologies have different advantages and disadvantages, so there is no universal data mining method. In the specific work, the appropriate algorithm is selected according to the data type and data set size.

However, in the process of data mining, an attribute is first determined to be included in the class attribute, and the number of determined attributes must be controlled, which is the concept of “multi-dimensional mining”. Because including too many attributes will reduce the performance of the system, then the system’s predictive responsiveness will be affected; If the included attributes are incomplete, then the accuracy of data mining will be reduced. Usually for a certain industry more familiar with the practitioners often through industry experience to filter out the relevant attributes. As the number of attributes in a dataset rises to a certain level, such as thousands of clinical indicators of diabetes, the difficulty factor for practitioners in making empirical judgments increases, resulting in misjudgments [11–13].

The indexes of clinical medical data mining model are shown in Table 2.

**Table 2.** Clinical medical data mining model indicators table

	First level indicators	Secondary indicators
Indicators of clinical medical data mining model	Basic attributes of patients	Age
		Sex
		Region
		Occupation
		Income
		Constellation
		Other indicators
	Medical services	Visit number
		Frequency of visits
		Consumption of medical treatment
		Other indicators
	Index business	Number of indicators
		Indication of abnormal indicators
		Index test value
		Index test frequency

Association rule algorithm is a frequent item set algorithm for mining association rules. Apriori algorithm and FP-Tree algorithm have been widely used in many fields such as sales, insurance and so on. The basic idea of the algorithm is to find out all item sets, the probability of occurrence of these item sets must not be less than the minimum support degree set by the system, and then generate strong association rules from item sets, which must not be less than the minimum support degree and the minimum confidence degree. Then, according to the rules generated by the initial itemset, all rules containing only the items of the set are generated, in which there is only one item on

the right of each rule, and only those rules that are greater than the minimum confidence level given by the system are left. The algorithm can generate all frequent item sets using a recursive approach.

In order to better explain the principle of association rule algorithm, this section starts with a simple medical data example, briefly describes the operation process of association rule algorithm. The sample data is shown in Table 3.

**Table 3.** Data sample table

Report sheet number	Medication status
1	A.B
2	C
3	D
4	A.B.C
5	A.C

Support: in the above sample, the support of sub item set  $\{A, B, C\}$  is  $1/5$ , because only the record with order number 4 contains all three sub items a, B and C. The support of  $\{A, B\}$  is  $2/5$ .

Confidence: this indicator will tell you how often a patient report contains a, then the possibility of item b. For example, the confidence level for  $\{A, B\}$  would be  $2/3$ , because three reports included a, and only two reports contained a sub item combination  $\{A, B\}$ . So the confidence is defined as:

$$Conf(X, Y) = \frac{Supp(X \cup Y)}{Supp(X)} \quad (1)$$

Promotion degree: in general, if the two have high support and high confidence, it means that there is a strong correlation between the two. However, in some cases, this may be incorrect, because even if A and B are independent of each other, items A and B may still have high support. A better criterion to judge the strength of association rules is called promotion degree. The calculation formula of lifting degree is as follows:

$$Lift(X, Y) = \frac{Supp(X \cup Y)}{Supp(X) * Supp(Y)} \quad (2)$$

The degree of promotion is greater than 1.0, which indicates that  $X, Y$  has strong correlation. The higher the degree of promotion, the stronger the correlation strength. In the previous example, the lift of  $\{A, B\}$  is calculated as follows:

$$Lift(A, B) = \frac{Supp(A \cup B)}{Supp(A) * Supp(B)} = 1.667 \quad (3)$$

### 3.3 Data Stability Control Module

This module uses the artificial intelligence control technology to carry on the accurate control to the clinical medical data stability, enhances the data monitoring the effect.

In the monitoring system of clinical medical data, there is a high demand for real-time data. This requires mobile platform -side applications to access the backend database at regular intervals, downloading the latest data in a timely manner. However, during the downloading of data, there are likely to be data loss, response timeouts, data redundancy, and so on. In the whole process of system, the system should specify the method of data time stability. In this paper, the use of frame drop management mechanism to complete. The frame drop mentioned here includes two cases: active frame drop and passive frame drop.

Because of the complexity of the network, the video packets transmitted in the network may be lost or damaged. If the received data is not checked by the receiver, then the decoder will be unable to decode normally. Network packets will contain a check code for the data they carry, in case the data is lost in the process of network transmission. This mechanism can detect the integrity of the data carried by a single packet, but if the entire packet is lost, it will be more difficult to detect.

This paper adopts SOAP protocol, which is a network protocol of sending packets sequentially. Then the serial number of the packet can be used to determine whether the received packet is lost. The loss of packets can lead to errors in data analysis, and once detected, the associated data needs to be discarded.

The design system has to discard some data that cannot be processed in time due to the limitation of the performance of the software and hardware of the device and the transmission capacity of the network. The behavior of frame dropping is initialized by the application program, so the case of frame dropping is called active frame dropping. Setting up active frame dropping is a control behavior that has to be made due to the special needs of monitoring. It needs to consider two aspects: when to lose frames and what data to lose. Handling active frame dropping requires control of at least three buffers located in different threads. The first thread is responsible for fetching packets from the network, hereinafter referred to as the fetch thread. It does not care about the contents of the packet and simply stores the obtained packet in the first buffer, hereinafter called packet buffer. The second thread parses the data in the packet into frames, hereinafter referred to as composite threads. The storage space set in this thread is called a combined buffer. The third thread is the frame data into the decoder for decoding, hereinafter referred to as waiting for the decoding thread. The data buffers in this are called waiting queues.

Through the design of the hardware unit and software module, the operation of clinical medical data monitoring system is realized, which provides an effective means for hospital administrators and public health management departments, and provides reliable data support for the development and research of clinical medical industry.

## 4 System Application Performance Analysis

In order to verify whether the design system can solve the problems existing in the existing system, this study uses MATLAB simulation platform to design experiments. The specific experimental process is as follows:

### 4.1 Construction of Experimental Environment

The experimental environment is the key to the smooth progress of the experiment, the specific building process as follows:

In order to ensure the stable operation of the system and the response efficiency of data analysis and intelligent supervision, the following hardware environment and corresponding system are planned.

The engine server mainly deploys the rules engine service. In a development environment, multiple engine services are started on a single engine server. Different engine services take different ports to provide communication. Scheduling and other application servers are mainly used for deploying ETL scheduling services, beforehand reminder services, Web application services and other application services.

Auditing engine is the core component of medical big data analysis and intelligent supervision system. The technical architecture of the audit engine is based on Microsoft's .NET Framework development framework. The audit engine is a relatively independent service, which needs to be run in a secure technical and software environment to ensure the security and uniqueness of data transmission and prevent confidential data from being leaked or tampered with. The .NET Framework is developed by Microsoft, a software development platform dedicated to agile software development, rapid application development, platform independence, and network transparency.

It is based on a programming platform which uses the system virtual machine to run, and supports the development of multiple languages based on the general language runtime library. .NET also provides new functionality and development tools for application programming interfaces. These innovations allow programmers to develop both Windows and Web applications as well as components and services. .NET provides a new reflective and object-oriented programming interface. .NET design so that many different high-level languages can be brought together. All languages in the .NET Framework provide a base class library.

This system server uses Oracle 11g as the database. Self-managing data: The ability of an Oracle 11g database to manage its own undo segments means that administrators no longer need to carefully plan and optimize the number and size of fallback segments, or worry about how to strategically allocate transactions to specific fallback segments. Memory management is another area of Oracle 11g where significant management is given. Because of the limitation of space, the management of database system improvement and simplification is not discussed too much.

### 4.2 System Function Setting

According to the requirements of clinical medical data monitoring experiment, the system functions are reasonably set to ensure the smooth progress of the experiment, as shown in Table 4 (Fig. 4).

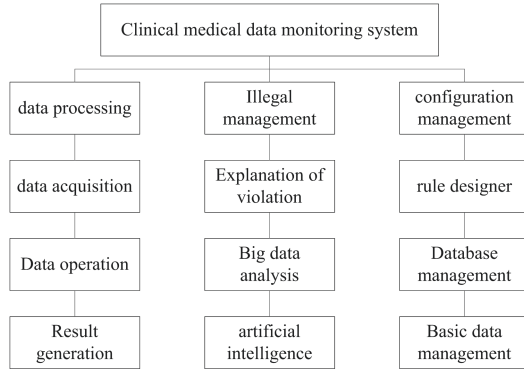


Fig. 4. System function setup diagram

### 4.3 Analysis of Experimental Results

Based on the abovementioned experimental environment and system functions, clinical medical data monitoring experiments shall be conducted to reflect the application performance of the system through system response time and monitoring efficiency. The specific experimental results are analyzed as follows:

The application performance data obtained from the experiments are shown in Table 4.

Table 4. Application performance data table

Number of experiments	Existing systems	Design systems
(1) System response time		
1	10.23 s	6.12 s
2	9.56 s	5.00 s
3	11.25 s	4.23 s
4	12.45 s	5.01 s
5	12.00 s	3.01 s
6	11.49 s	4.09 s
(2) System monitoring efficiency		
1	45.23%	70.12%
2	56.59%	75.49%
3	58.40%	80.00%
4	60.12%	81.24%
5	51.47%	80.05%
6	52.05%	85.49%



As shown in Table 4, the existing system response time range is 9.56 s–12.45 s, and the monitoring efficiency range is 45.23%–60.12%; the design system response time range is 3.01–6.12 s, and the monitoring efficiency range is 70.12%–85.49%. The main reason is that in the design of the system, big data technology is used to mine and preprocess clinical data, which improves the performance of the system. Compared with the existing system, the response time of the design system is shorter and the monitoring efficiency is higher, which fully proves that the design system has better application performance.

## 5 Conclusion

By introducing artificial intelligence and big data analysis method, a new clinical medical data monitoring system is designed, which greatly shortens the system response time and improves the system monitoring efficiency.

## References

1. Ho, C., Joseph, A., Karel, C.: Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance. *Bull. World Health Organ.* **98**(4), 263–269 (2020)
2. Kotsenas, A.L., Balthazar, P., Andrews, D., et al.: Rethinking patient consent in the era of artificial intelligence and big data. *J. Am. Coll. Radiol.* **18**(1), 180–184 (2021)
3. Cheng, X., Chen, F., Xie, D., et al.: Design of a secure medical data sharing scheme based on blockchain. *J. Med. Syst.* **44**(2), 52 (2020)
4. Li, B., Ding, S., Song, G., et al.: Computer-aided diagnosis and clinical trials of cardiovascular diseases based on artificial intelligence technologies for risk-early warning model. *J. Med. Syst.* **43**(7), 228 (2019)
5. Goldenberg, S.L., Nir, G., Salcudean, S.E.: A new era: artificial intelligence and machine learning in prostate cancer. *Nat. Rev. Urol.* **16**(7), 1 (2019)
6. Kalinin, S.V., Lupini, A.R., Dyck, O., et al.: Lab on a beam—big data and artificial intelligence in scanning transmission electron microscopy. *MRS Bull.* **44**(7), 565–575 (2019)
7. Xu, K., Wang, Z., Zhou, Z., et al.: Design of industrial internet of things system based on machine learning and artificial intelligence technology. *J. Intell. Fuzzy Syst.* **40**(2), 2601–2611 (2021)
8. Dong, D., Wang, X.: Human-computer system design of entrepreneurship education based on artificial intelligence and image feature retrieval. *J. Intell. Fuzzy Syst.* **39**(4), 5927–5939 (2020)
9. Ding, Y.: Performance analysis of public management teaching practice training based on artificial intelligence technology. *J. Intell. Fuzzy Syst.* **40**(5), 1–14 (2020)
10. Foresti, R., Rossi, S., Magnani, M., et al.: Smart society and artificial intelligence: big data scheduling and the global standard method applied to smart maintenance. *Engineering* **6**(7), 835–846 (2020)
11. Liu, S., Bai, W., Liu, G., et al.: Parallel fractal compression method for big video data. *Complexity* **2018**, 2016976 (2018)

12. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. *Multimedia Tools Appl.* **76**(4), 5787–5802 (2014). <https://doi.org/10.1007/s11042-014-2408-1>
13. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–17 (2018). <https://doi.org/10.1007/s11036-018-1134-8>



# Construction of Pediatric Medication Data Security Cloud Storage Model Based on Internet of Things Technology

Shu-hua Whang<sup>1</sup>, Wen-shan Yao<sup>1</sup>, Xian-ying Meng<sup>2</sup>, Min-na Zheng<sup>1</sup>,  
and Hua Fan<sup>1</sup>(✉)

<sup>1</sup> Medical and Nursing Branch, Panjin Vocational and Technical College,  
Panjin 124010, China

{zmfds222, jwqdds2000}@aliyun.com

<sup>2</sup> Department of Foundation, Panjin Vocational and Technical College,  
Panjin 124010, China

**Abstract.** According to the principle of database modeling and design standardization, combined with the design requirements of pediatric medication information database, the construction method of pediatric medication data security cloud storage model based on Internet of things technology is proposed, the overall structure of pediatric medication information model database is designed, and the overall planning of data storage model, the overall design and implementation of model database are given, the Internet of things technology is selected to optimize and describe the security structure of the database model, and the performance optimization strategy of the pediatric medication information model database is proposed, in order to lay the foundation for the construction of a multi-layer architecture and high efficiency medication data security cloud storage model.

**Keywords:** Internet of things technology · Pediatric medication · Secure storage

## 1 Introduction

At present, there are many problems in the circulation of paediatric drugs, such as the change of environment will lead to the change of drug quality or even complete failure. In the circulation process, a large number of fake drugs may be mixed, goods are in collusion and return. The virtual increase of circulation links. How to manage pediatric drugs effectively is a model project, involving many aspects, which has attracted widespread public attention [1]. In order to improve the efficiency of pediatric medication data management and the security of data storage, it is necessary to conduct research on the safe cloud storage method of pediatric medication data. This method utilizes modern means such as computer hardware and software technology and network communication technology to comprehensively manage pediatric medication information. It collects, stores, processes, extracts, transmits, summarizes and processes the data generated in each stage of medical activities to generate various information,

so as to provide a comprehensive and automatic management and information management system of various services for the overall operation of the hospital [2].

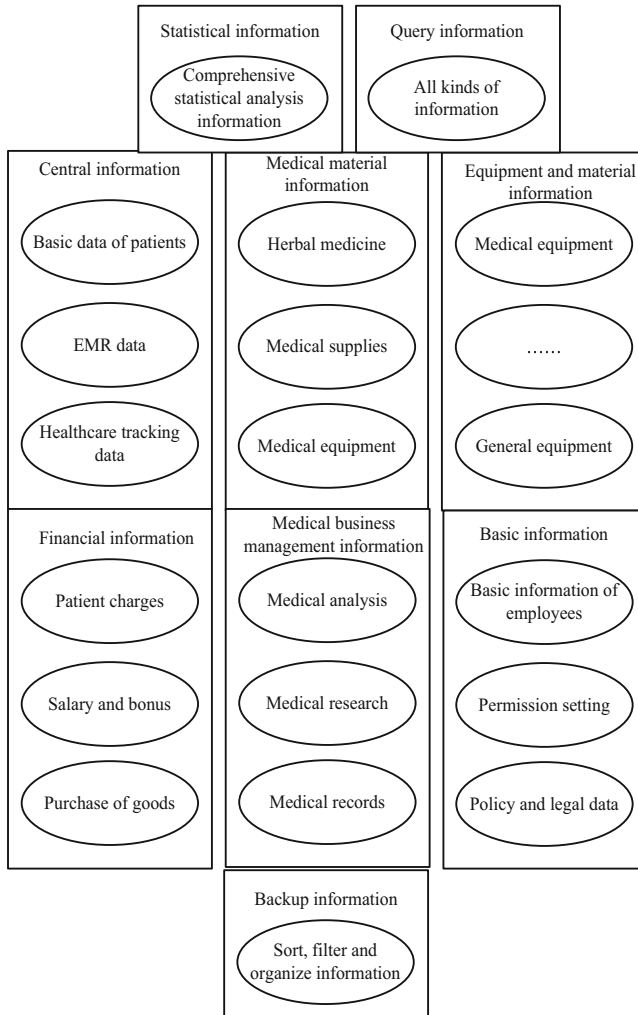
At present, the study of the method for data security cloud storage has a lot of, for example, reference [3] proposed an identity based encryption the data security of cloud storage method, this method is mainly aimed at the demand of the users to share the data security in cloud storage, the identity of the proposed hierarchical encryption, identity management, improve the efficiency of the private key generation and the security of key management. The method of private key version number is introduced to improve the CP-ABE algorithm and improve the efficiency of attribute retraction, so as to further enhance the security of data cloud storage. Reference [4] proposed a data secure cloud storage method based on explicit exact minimum storage regenerated code. This method uses EEMSR code to encode the data, and then uploads the encoded data to the cloud to ensure the availability of the data. The encrypted hash function can verify the integrity of the cloud data through the Challenge-Response protocol. The EEMSR code is a regenerative code that can accurately recover lost chunks of data with less repair traffic, both of which enable secure cloud storage of data.

However, the above methods have poor adaptability in the safe cloud storage of pediatric medication data. Therefore, this paper builds a safe cloud storage model of pediatric medication data based on the Internet of Things technology. In order to solve the specific problems in the operation of pediatric medication information model and improve the security of pediatric medication data storage, it has a very important contribution and significance for the further development of medical information security field.

## 2 Pediatric Medication Data Security Cloud Storage Model

### 2.1 Design of Pediatric Medication Database

Pediatric drug use information model is a large online transaction processing model. There are many kinds of information in the model, large amount of data saved and historical data accumulated. This will lead to the increasing space occupation of the model and the corresponding decrease of the efficiency of the model. Therefore, in the overall design stage, we must plan the storage model of business data from a long-term perspective. The principle of master plan is to control the size of daily operation data table under the premise of ensuring data security. Therefore, the following two measures are taken: to establish logical database, which can be classified, stored and kept, and physical distributed database can be realized according to the needs; The overdue data can be classified, screened and sorted, stored regularly and separated from daily operation data tables. On the one hand, the model efficiency can be improved and the old data query will be facilitated [5]. The overall structure of the pediatric drug use information model database is shown in Fig. 1 below.



**Fig. 1.** Structure of pediatric medication database

By attaching labels on drug packaging and installing sensor devices in all places where drugs pass through, the identification and real-time acquisition of basic attribute information, status information and process information of drugs can be realized, and the collected and perceived data can be stored and transmitted in the form of Internet of things standard language PML, so as to make drug manufacturers drug wholesale companies, logistics companies, retail pharmacies, hospitals, drug buyers, superior authorities and other entities can obtain the information they care about in real time and transparently through cable [6]. In order to facilitate understanding, based on the knowledge of set theory, the physical meaning of each element in the drug circulation IOT model is shown in Table 1 below.

**Table 1.** Physical meaning of elements in IOT model of drug circulation

Describe	Physical significance and description
Input of drug circulation model	RFID tag number of drug entity object
Configuration document of drug circulation model	It is used to store parameters related to the description of drug circulation information
Output of drug circulation model	The output information includes: authenticity of drugs, basic attribute information of drugs, status information of drug circulation at a certain time, generated business documents and process information of drugs in the whole circulation
Storage template of drug circulation information	Define the standard structure of drug circulation information storage through PML header

The pediatric drug entity object is defined as the object set  $S = \{S_1, S_2, \dots, S_n\}$  and  $S$ , which can be a single drug or a package box or box composed of a batch of single products  $S = \{S_1, S_2, \dots, S_n\}$ . Correspondingly, the object set  $S$  passes through a first-class point at time  $t$ , and its state information  $Pt(s)$  can be described by nine elements.

$$P_t(S) = |Attr(S), t, Id(S), Loc(S), Envi(S), User(S), Task(S), Rela(S), Ord, S| \tag{1}$$

The specific meaning of each element is shown in Table 2.

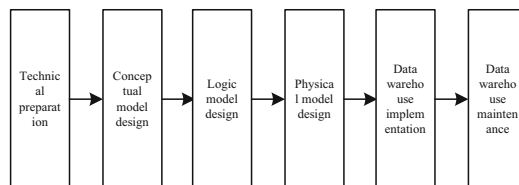
**Table 2.** Characteristic description of drug status information

Element	Describe	Meaning and explanation
Id(S)	RFID tag number of drug entity object S	Because the drug entity object s can be a single drug, a packing box or a packing box, its corresponding label number is also different
Envi(S)	Environmental information of the location of the current drug entity object S	Through the sensor of drug storage location
Attr(S)	Basic attribute values of drug entity object S	Provided by drug manufacturers, such as drug composition, properties, pharmacology, etc.
t	Time when the state of drug entity object S changes	Automatic generation of S-tag of drug entity object read by RFID reader
Loc(S)	The place where the status of drug entity object s changes	The location LOC information can be obtained by reading the RFID reader number
Rela(S)	The RFID tag number of the entity object related to the drug entity object S	It is mainly used to express the relationship between a single drug and the packing box when the drug retail is unpacked
User(S)	The ID number of the operation user that causes the change of the s state of the drug entity object	They can be drug manufacturers, drug wholesalers, logistics companies, drug retailers and buyers
Ord(S)	The business document number corresponding to the drug entity object S	This variable is used to realize the association with the business data of the enterprise entities in circulation
Task(S)	Business type of drug entity object s in circulation process	The business types here mainly include packing case warehousing, packing case outbound, transportation, unpacking and retail

Based on the characteristic information in the above table, the circulation process of pediatric medication is described.

$$\text{Proc}(\text{Id}(S)) = \{\text{Attr}(S), \text{Loc}(S), \text{User}(S), \text{Task}(S), \text{Ord}(S)\} \quad (2)$$

Through the RFID number of electronic label of paediatric drug entity  $S$ , we can query the authenticity of paediatric drug use, basic attribute information of drug and the whole process information in pediatric drug circulation  $S$  [7]. When storing and expressing pediatric drug use information, users with different authority care about different circulation information, such as ordinary consumers are more concerned about authenticity, basic attribute information and environmental problems of circulation place of purchased drugs, while quality supervision department pays more attention to the inquiry of pediatric drug flow process, so that when problems occur in pediatric medication, drugs can be recalled in time, minimize the hazard. Therefore, in order to facilitate the real-time query of circulation information of various users, a storage model of pediatric drug use information based on PML is constructed [8]. Data warehouse is an integrated, subject-oriented, time-varying and stable data collection. The main function of data warehouse is to improve the efficiency of decision-making and analysis, and reorganize the corresponding data according to the application needs of specific fields, so as to produce a data environment that can adapt to decision-making [9]. Compared with the traditional database, data warehouse has a great difference. Data warehouse is mainly a means of obtaining information from massive data. The research results can be used to provide effective decision support for the decision-making of management. The traditional pediatric drug use database model is applied, detailed, small amount of primary operation data, can be updated, and is mainly used in the daily data operation process of database model. It is a transaction oriented design, and the corresponding data is stored in an operational database. However, data warehouse is an analytical, refined, one-time operation data with large amount and can not be updated, which is mainly used to manage data requirements and requirements in decision analysis. The data is stored in an analytical database. With the data abstraction tool, the data of pediatric medication which is useful for this model can be extracted from the mass information. After the integration and reorganizing, the data environment can be generated to adapt to the decision-making, which provides effective data support for decision makers to formulate corresponding specific policies. The data of pediatric medication is the theme of this data warehouse. Usually, the establishment of a data warehouse can be completed only according to the following steps (Fig. 2).



**Fig. 2.** Steps of pediatric medication information management

Through the standardized modeling of the basic attribute information, operation user information, business type information, business document information, circulation status information and circulation process information of drugs, the relevant information of drug circulation can be effectively stored and transmitted.

## 2.2 Data Security Algorithm for Pediatric Drug Network

In order to fully understand the information requirements of the Internet of things for drug circulation in tracking and tracing, it is necessary to model and analyze the tracking process of pediatric drug circulation, so as to plan the development of application platform through the model [10]. As the Internet of things has a graphical development interface and rigorous mathematical basis, it can play a normative role in guiding the development of the application platform in the later stage, so the Internet of things is selected to model and analyze the pediatric medication management information [11]. Using the incidence matrix method, the Internet of things model of drug information tracking process and the relevant definitions of the incidence matrix are as follows:

$$\left[ C_{ij}^- \right]_{1+ \times 16} = \begin{cases} 1, & i = j \in [0, 1] \\ 1, & j = i + 1, i \in \{2, 3, \dots, 13\} \\ 0, & \text{other} \end{cases} \quad (3)$$

$$\left[ C_{ij}^+ \right]_{14 \times 16} = \begin{cases} 1, & j = i + 1, i \in [0, 1] \\ 1, & i = 1, j = 3 \\ 1, & j = i + 2, i \in [2, 3, \dots, 13] \\ 0, & \text{other} \end{cases} \quad (4)$$

In the process of management of pediatric drug storage model, two considerations are mainly based on: one is the level of user service, that is, there are enough suitable goods in the right place and time. The other is the order cost and the storage model holding cost to minimize the sum of the two. It can be seen that the management of pediatric drug storage model takes a single enterprise as the object, and its main purpose is to control the storage model level of the enterprise on the premise of meeting the customer service requirements, strive to reduce the storage model level as much as possible, improve the efficiency of logistics model, and strengthen the competitiveness. The storage model is classified and managed, and the ordering time and quantity are determined to minimize the total cost of the storage model. The storage model management decision model is designed and analyzed from the perspective of determining the quantity of each purchase. Its purpose is to determine the purchase quantity of storage model, so as to reduce the cost of storage model [12]. The establishment process of storage model is as follows:

- (1) Query the historical data of sales volume and storage model volume of various pediatric drugs, and the specific data of current storage model pediatric drugs, and determine the data source of the model. The storage model commodities are classified and managed according to ABC classification.
- (2) Referring to the previous pediatric drug purchase data, considering the constraints of funds and storage model capacity, the external factors such as demand,



customer service level and safety inventory that affect the purchase expectation were reasonably predicted, and different parameter values were set to determine the external constraints.

- (3) According to the established pediatric medicine comprehensive storage model, the optimal pediatric medicine ordering quantity of main storage model materials under the optimal total cost was obtained [3].
- (4) The results provide decision-making reference for senior managers. If the results meet the requirements, they can be executed. Otherwise, users can increase their own judgment on the basis of the output results of the reference model, so as to make a more reasonable storage model decision-making scheme, which greatly increases the flexibility of decision-making.

There should be pharmacy storage model information in pediatric medication database, and pharmacy storage model should indicate the pharmacy information. Pediatric drug use data storage model needs to know the associated drug SKU, while drug SKU does not need to know the storage model information. In addition, there are several groups of combination relationships in the pediatric medication database model, such as order class and order sub item class. One or more order sub items constitute an order. It is meaningless to talk about order sub items alone. When an order is created, the order sub items are created at the same time. When an order is deleted, all the same order sub items are deleted. After combing and analyzing these association relations, we also sort out the rank mapping relations with association relations, based on which we can get the association information of pediatric medication data category characteristics, and finally record the sorted information in the Table 3.

**Table 3.** Association of pediatric medication data category characteristics

Association class	Relationship	Order element relation
Drug delivery task-robot arm	Two way association	0 or 1 to 1
Drug SKU-order sub item	One way association	One to many
Pharmacy-pharmacy inventory	Two way association	One on one
Drug SKU -warehousing record	One way association	One to many
Drug SKU-drug delivery subtask	One way association	One to many
Drug SKU-drug inventory	One way association	One to many
Drug SKU-outbound record	One way association	One to many
Drug SKU-frozen inventory records	One way association	One to many
Pharmacy-orders	One way association	One to many
User- order	One way association	One to many
Medicine-medicine SKU	One way association	One to many
Dispensing task-dispensing window	One way association	0 or 1 to 1
Pharmacy-robotic arm	Synthetic relation	One to many
Pharmacy inventory-warehousing records	Synthetic relation	One to many

Based on the above table management information, the decision-making process of pediatric medication database management was further optimized, as shown in Fig. 3.

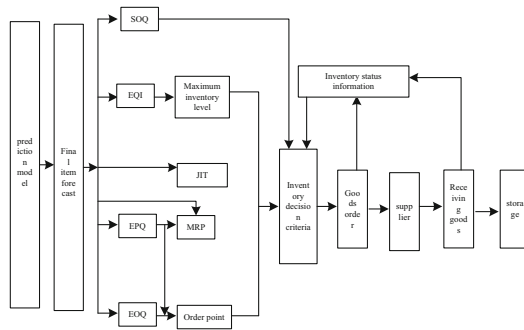


Fig. 3. Framework of drug storage management decision process

According to the characteristics of drugs, the statistical index with practical significance is selected as  $P$ ,  $Q$  reflects the value of stored drugs and the level of use value, then the basic statistical index algorithm is.

$$W_j = Q_j \cdot P_j \tag{5}$$

Assuming that  $q_j$  is the intact rate of the storage model, the degree of difficulty or possible damage or deterioration of children’s medication is further calculated

$$u = \lfloor (Q_j - q_j) / W_j + Q_j \rfloor \times 100\% \tag{6}$$

After the classification of storage models, different levels of storage models are managed according to the drug category, so as to selectively control the storage models and reduce the pressure of storage model management. Pediatric drug market forecasting process can be regarded as a guidance model, based on which the pediatric drug demand forecasting process is optimized, and the specific structure is shown in Fig. 4.

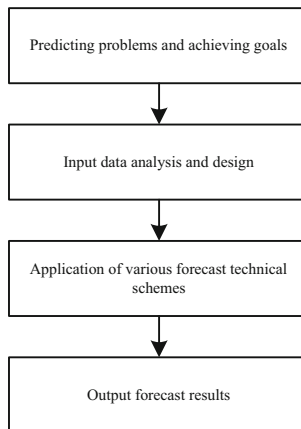


Fig. 4. Diagram of pediatric drug demand prediction process

As shown in the figure, the reliability of pediatric drug information processing can be improved by collecting and analyzing historical data.

### 2.3 Construction of Pediatric Medication Data Security Storage Model

According to the nature of pediatric medication data storage methods, they can be divided into qualitative prediction and quantitative prediction. Qualitative market prediction is a subjective prediction based on the nature and regulation of things. Quantitative prediction is based on the historical data and related factors of things, using statistical methods and mathematical models to study and speculate the use of pediatric drugs and its structural relationship, and predict various trends of production, sales and market demand. The model is divided into two modules: drug storage early warning module and intelligent decision support module (Tables 4 and 5).

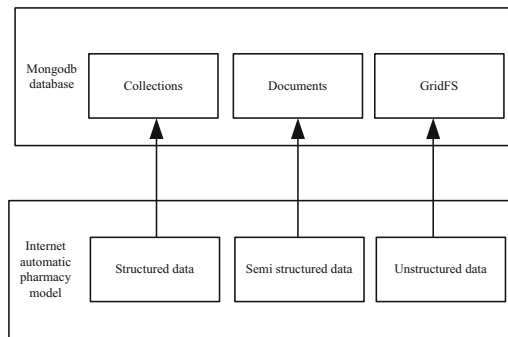
**Table 4.** Function list of drug storage model management module

Functional module	Corresponding function name
Basic information management	Model information settings
	Drug dictionary editor
	Model login
	Drug base setting
Statistical inventory management	Inventory report
	Inventory forecast report
	Inventory alert report
Operator management	Management operator
Drug inventory management	Forecast model management
	Inventory model management
	Lower limit setting of drug inventory
	Upper limit setting of drug inventory

**Table 5.** Function list of intelligent decision support module

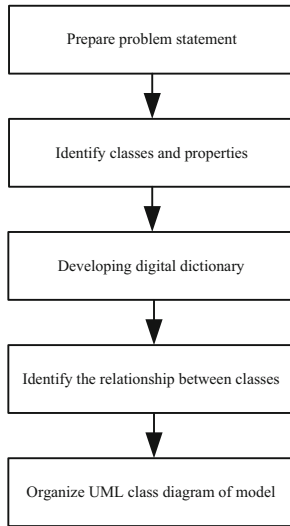
Functional module	Corresponding function name	
Decision support	Selection of inventory forecasting model	Weighted average forecasting model
		Seasonal forecasting model
	Inventory management model selection	Discrete storage model
		Continuous storage model
		Model of instant purchase and allowed shortage
		Continuous purchase and allowed shortage model
		Instant purchase, price discount
		Model of instant purchase and no shortage
		Continuous purchase and no shortage model

In the process of model design, the quantitative market forecasting method is selected in order to make the forecasting results as accurate and objective as possible. There are many kinds of forecasting contents for both qualitative and quantitative methods. Therefore, considering the design emphasis, only the storage model demand forecasting is described quantitatively. Using mongodb database to design the storage scheme of structured, semi-structured and unstructured data in the model. The structured data in the model will be stored by the database design method. Because the storage form of mongodb database is different from the traditional relational data storage model, the characteristics of mongodb database need to be considered in the database design. The electronic prescription and other document information are stored in Mongo DB documents through certain transformation. The images, videos and other data in the model will be stored by the Grid FS mechanism provided by MongoDB. Based on this, the storage mode of children's drug information is optimized, as shown in Fig. 5.



**Fig. 5.** Storage mode of children's drug information

The domain modeling idea is used to complete the requirement analysis of the model data, and the UML class diagram is used to complete the model design. The first step of model database design and development is to understand the function of the model and analyze its requirements. The first is the understanding of the domain model, that is, the understanding of the problem domain of the real model. Requirements analysis needs to find out the general requirements of the domain model, that is, the general requirements, and then organize the general requirements and description into a standardized problem statement. And the problem statement is analyzed, modified and improved. Finally, the class diagram model and data dictionary are constructed by object analysis technology. The whole requirement sorting process is shown in the figure, which is divided into five stages, namely, preparing problem statement, identifying class and attribute, developing digital dictionary, identifying association between classes, sorting UML class diagram of model, and constructing static domain data modeling steps of pediatric medication, as shown in Fig. 6.



**Fig. 6.** Static domain data modeling steps

Demand forecasting is mainly to forecast the future sales volume of goods, in order to provide the necessary demand parameters for the storage model decision. At present, there are many forecasting models. Which forecasting model an enterprise chooses depends on many factors, such as the acquisition of historical data, forecasting preparation time, forecasting cycle, forecasting accuracy, forecasting cost, and product characteristics. Other issues that need to be further considered are the degree of flexibility of the enterprise (the stronger the enterprise's ability to respond quickly to changes, the higher the enterprise's flexibility), the lower the accuracy of the prediction model, and the consequences of bad prediction. The model of using inventory to meet the demand in time is a meaningful service index. The service level coefficient of demand can be defined as:

Service level coefficient of demand = Supply/Total demand.

Shortage level coefficient of demand = Shortage/Total demand.

The above relationship must be measured for a certain period, which can be the duration of lead time. The expected shortage quantity in an order cycle is:

$$E(M > B) = \int_B^0 (M - B)f(M)dM \quad (7)$$

To get the out of stock level factor of the demand quantity in the order period, it is necessary to divide the demand quantity within the order period.

$$W = \frac{E(M > B)}{Q} \quad (8)$$

At present, there is still no corresponding decision support function for children's medication model, especially in the process of children's medication storage model management, decision support is one of the important issues concerned by the medical insurance department. At present, children's medication needs another function that storage model management model can provide, that is, it can provide data decision support for children's medication storage model management. All nouns and noun phrases are extracted from the problem statement, and the meanings of the extracted nouns and noun phrases are carefully considered, and whether they may become classes or attributes is screened. Based on this, the key words structure of children's medication attributes is standardized as follows (Table 6).

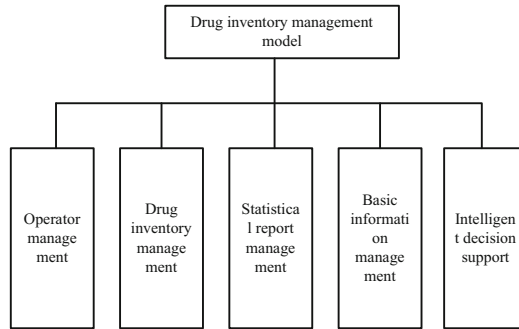
**Table 6.** Key words of medication attributes for children

User (role playing)	Drugs (definite things)
Prescription name (simple value, attribute)	Login (event)
Prescription address (simple value, attribute)	Drug freezing mechanism (concept)
Pharmacy for taking medicine	Mechanical arm (definite thing)
Take medicine window (clear things)	Window number (simple value, property)
Operation indication (simple value, attribute)	Pharmacy status (simple values, attributes)
Order (concept thing)	Drug quantity (simple value, attribute)
Drug delivery task (concept)	Manipulator status (simple values, attributes)

Generally, the management department of drug storage model mainly needs to obtain the following decision support information:

- (1) The storage model management department needs to determine which storage model to choose, or adjust the storage model management model as needed.
- (2) The storage model management department needs to count and analyze the warehousing data information in detail to understand the warehousing demand of drugs in different situations, so as to set a reasonable storage model lower limit.
- (3) The storage model management department needs to count and analyze the data in detail, so as to understand the quantity of drugs in different cases, and then set a reasonable upper limit of the storage model.
- (4) Reasonable choice of storage model and prediction model.

Based on the above steps, the general framework of children's medication storage model management is optimized, as shown in Fig. 7.



**Fig. 7.** Framework of drug data storage model management system

It can be seen from the figure that the data storage model management model for children's medication mainly includes five parts:

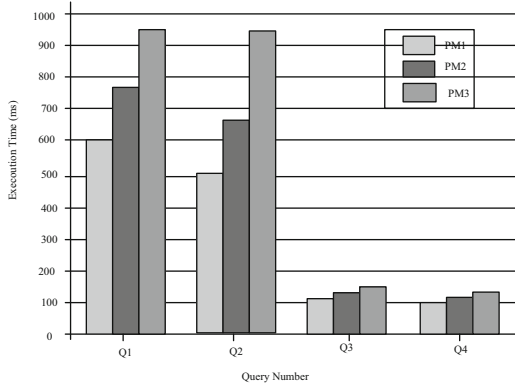
- (1) Basic information management: including drug base setting, drug dictionary editing, etc.
- (2) Drug storage model management: mainly includes drug storage management and drug delivery management.
- (3) Statistical report management: mainly used for the management of drugs, financial and other statements.
- (4) Operator management: mainly used for the management of operators.
- (5) Intelligent decision support: mainly used for storage model alarm, model selection and prediction.

This model needs to count the demand and supply of drugs in hospitals or pharmacies, such as the total category of drugs, the number of suitable people, the level of drug safety, etc., so as to reasonably store or issue drugs according to the total amount of current drug storage model. Data mining methods such as analysis of variance, multiple linear regression and multi-element analysis are used to provide alarm information. In order to avoid the shortage of drug supply or a large number of drug waste and other problems, so as to manage the reasonable and effective drug storage model. The storage model management model should be effectively determined or timely adjusted according to the current operation situation, or fixed and reasonable storage model management model should be selected according to the needs. These models include deterministic model and stochastic model. To provide effective decision support for the storage model management, to meet the demand of drug supply, does not cause a waste of resources.

### 3 Analysis of Experimental Results

The specific functions of the model are verified, and the detailed contents of the model function design are given, including each module and the specific related functions under each module. The experimental results of some specific functions of the model are shown.

In the experiment, four computers in the lab LAN are used to build a distributed MongoDB cluster. Structured data and unstructured documents are stored in the Collection of MongoDB database in JSON format. The communication bandwidth of each machine is 10 Mbps, the processor is Lnter (R) Core (TM) i5 CPU, 2.40 Hz, memory is 4 GB, hard disk is 320 GB, and Java 1.7.0 is used to implement the index scheme proposed in this chapter. We build test data for the medical field. The structured data set is a database about drugs collected from a medical open website, which contains about 8000 drug records. Then, a web search was performed on the PubMed search engine of the children's medication database with each medication name as a keyword, and the first five items of the results were used as unstructured test data set. Three test data sets with different sizes were constructed, namely 350 MB(PM1), 700 MB(PM2) and 1G (PM3). For each dataset, a process with a concurrency of 4 is used to create a data security index for the mixed data. Based on this, the information query effect of pediatric medication data security storage model is tested, and the experimental results are shown in Fig. 8.

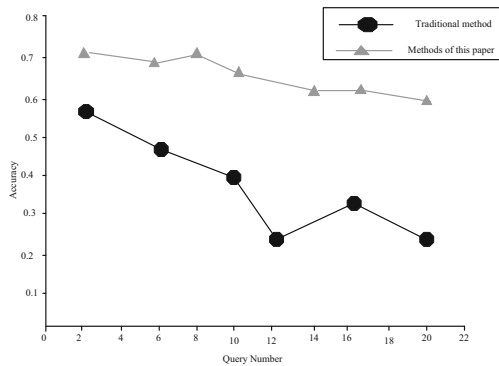


**Fig. 8.** Execution result of drug storage information query

It can be seen from the figure that when the amount of data gradually increases, the execution time of each query increases slightly, but it can be completed at the second level. With the same amount of data, the execution time of entity queries Q1 and Q2 is significantly less than that of relational queries Q3 and Q4. This is because relational queries also involve the semantic relationship between entities, so the execution process is relatively complex and needs more time. In addition, it can be seen from the figure that for the same type of queries, such as entity queries Q1 and Q2, the query



execution time increases only slightly with the increase of query complexity, indicating that the index mechanism has good scalability in processing the same type of queries. In the experiment, we compare the association index mechanism with the independent index mechanism, so as to verify the advantages of this scheme. The independent index mechanism uses two different indexes for structured and unstructured data, that is, the traditional document oriented inverted index and the database oriented Bree index. The two indexes do not interfere with each other, and the query is processed based on keywords. For the test data, we construct 10 queries manually, including both entity query and relation query. Then, the associated index and independent index mechanism are used to execute the 10 queries, and the accuracy of the query results is calculated. The accuracy is obtained by dividing the number of documents and records related to the query object in the query result set by the total entries in the query result set. The comparison results are shown in Fig. 9.



**Fig. 9.** Accuracy of pediatric medication data storage query

Analysis above, the pediatric drug use data stored query precision of this method is always higher than the traditional method, the reason is that the method of pediatric medicine information model database design, the overall structure of data storage model is also given the overall planning, overall design and implementation of model database, choose the Internet of things technology of database security model structure is optimized and description, Therefore, it has high precision of pediatric medication data storage and query.

## 4 Concluding Remarks

This paper mainly designs the pediatric medication data security cloud storage model based on Internet of things technology. The construction of hospital information model should follow the principles of meeting business needs, being practical and solving problems. As the background support of HIS, the model database was applied in the primary pediatric medicine. The results show that the pediatric medicine database

model is easy to operate, and all kinds of documents meet the actual requirements of the hospital, reaching the pre design standard. The security optimization of pediatric medicine database performance is realized, and the specific problems in the process of data security cloud storage model operation are well solved.

## References

1. Khan, F., Rehman, A.U., Yahya, A., et al.: A quality of service-aware secured communication scheme for internet of things-based networks. *Sensors* **19**(19), 4321–4327 (2019)
2. Li, W.: Design of smart campus management system based on internet of things technology. *J. Intell. Fuzzy Syst.* **40**(2), 3159–3168 (2021)
3. Zeng, D.S., Luo, J.W., Gao, J., et al.: Design of data security sharing model in cloud storage environment. *Intell. Comput. Appl.* **9**(4), 240–247 (2019)
4. Yang, J., Tan, D.J., Shao, J.X.: Research on the security of stored data in cloud computing. *J. Chongqing Univ. Posts Telecommun. (Nat. Sci. Ed.)* **31**(5), 710–715 (2019)
5. Wu, M., Hong, L., Zhao, Y., et al.: Dosage prediction in pediatric medication leveraging prescription big data. *IEEE Access* **1** (2019)
6. Liu, T., Liu, X., Li, X., et al.: Cloud enabled robust authenticated key agreement scheme for telecare medical information system. *Connect. Sci.* **12**(13), 1–20 (2021)
7. Milenkovic, A., Jankovic, D., Rajkovic, P.: Extensions and adaptations of existing medical information system in order to reduce social contacts during COVID-19 pandemic. *Int. J. Med. Inform.* **141**(9), 104224 (2020)
8. Domingues, M., Rui, C., Rodrigues, P.P.: CMIID: a comprehensive medical information identifier for clinical search harmonization in Data Safe Havens. *J. Biomed. Inform.* **114**(1), 103669 (2020)
9. Chen, H.Y., Wu, Z.Y., Chen, T.L., et al.: Security privacy and policy for cryptographic based electronic medical information system. *Sensors* **21**(3), 713 (2021)
10. Du, M., Chen, Q., Chen, J., et al.: An optimized consortium blockchain for medical information sharing. *IEEE Trans. Eng. Manag.* 1–13 (2020)
11. Zhang, L.: Research on storage data security of network service cloud platform based on OSI model. *Mod. Electron. Tech.* **43**(5), 74–77 (2020)
12. Liu, X.J., Ye, W., Jiang, J.W., et al.: Secure data storage scheme in hybrid cloud. *Trans. Beijing Inst. Technol.* **39**(03), 295–303 (2019)



# Mathematical Model of Data Partition Storage in Network Center Based on Blockchain

Bing-bing Han<sup>1</sup>, Zai-xing Su<sup>1</sup>, and Hai-yun Han<sup>2</sup>(✉)

<sup>1</sup> Panjin Vocational and Technical College, Panjin 124000, China  
dnsdsdf21@aliyun.com

<sup>2</sup> School of Humanities and Social Sciences, Sanya Aviation & Tourism College, Sanya 572000, China  
sdgsahd1232@aliyun.com

**Abstract.** In view of the poor effect of data partition storage in current network center, which affects the effect of data storage, this paper proposes a mathematical model of data partition storage in Network Center Based on blockchain. Firstly, the cross attribute storage algorithm based on blockchain is designed. Through the algorithm, the function of network center data is analyzed, and different functions are analyzed in detail. On this basis, the method of network center data division is designed. Finally, the network center data is divided into two parts. Build the mathematical model of data partition storage in network center to improve the cache performance. The experimental results show that the proposed method can better meet the requirements of the central data partition storage mathematical model cache operation efficiency, and provide better storage and operation performance.

**Keywords:** Blockchain · Network center · Central data · Data storage · Mathematical model

## 1 Introduction

Today's Internet applications pay more and more attention to the access efficiency of data partition storage in the network center. In order to better improve the effect of data storage, we must maximize the use of processor for partition processing, and store the data that may be used in multi-level buffer [1].

Based on this, this paper proposes the construction of the mathematical model of network center data partition storage based on blockchain, so as to better in the process of partition storage, the location of data placement is particularly important for the optimization of cache utilization. Therefore, in the operation of data partition, we should choose a good data storage scheme, improve the spatial layout of data distribution, improve the utilization of cache, and improve the performance. The storage layout of each attribute value in the record is adjusted, and some attributes in the record are accessed according to the requirements, so as to eliminate the memory delay caused by unnecessary memory access.

## 2 Mathematical Model of Data Partition Storage in Network Center

### 2.1 Data Feature Management in Network Center

With the rapid increase of data volume, network center data feature management has been updating. Static data is the core of data partition center in Network Center, and the feature of static data is metadata, so the collection of metadata is a key. In the aspect of metadata collection, the most traditional way of collection management is to store the program and data together, which brings many restrictions to the flexibility of the system, so many data and programs have begun to collect and store separately [2]. Until the emergence of database feature management information, the separation of metadata and basic data is realized. Metadata management is mainly in two aspects. One is the storage of metadata, which serves as the access docking and storage place of metadata. The second is a way of metadata exchange. Currently, there are three main metadata management models, such as Table 1:

**Table 1.** Data management model of Network Center

Type	Explain	Advantage	Shortcoming
Centralized model	There is a metadata server that serves the metadata store with the client request	Simple implementation	Single point of failure, limited storage
Distributed model	With the cooperation of multiple hosts, it serves as the metadata management center of the whole system	Solve some single point problems	Metadata synchronization is expensive and difficult to design
No metadata model	Using hash algorithm to replace metadata service in distributed system	Eliminate performance bottlenecks, single point of failure, etc.	Increase data management complexity

In the aspect of distributed database, the distributed database can be regarded as a whole in the logical layer and a distributed data system in the physical layer. Distributed database has unique advantages in applicability, reliability, availability and scalability. Before the tide of Internet and informatization changes the demand of database, we are faced with a small number of users. The traditional relational database can fully cope with it, and the distributed database is put aside because of its consistency problem [3]. However, in the face of a huge amount of data, the traditional relational database technology has been unable to meet the requirements of massive data. As a result, the requirement of data consistency is lowered, and NoSQL database, which is often used now, appears. Column storage and traditional row storage are two corresponding storage methods [4]. In many practical scenarios, what is needed is only an individual field of a record, and too many queries are unnecessary, especially in the massive data environment, the cost can not be underestimated [5]. For the applicable

column storage, the query is grouped according to the “column”, and the query is concentrated on the corresponding column, which makes the operation of individual attributes more convenient. The current mainstream column storage database is shown in the Table 2.

**Table 2.** Data sequence of network partition

Mainstream representatives	Hbase, Cassandra, Hypertable
Data model	Centralized data storage by column
Advantage	For column query, IO has obvious advantages and strong scalability
Shortcoming	Data integrity is no better than row storage

## 2.2 Data Partition Method in Network Center

Data storage is related to the use of data, performance, and management difficulty. Data quality evaluation is related to user experience, resource allocation, and accuracy of results. Data storage and quality evaluation are very important for the efficient use of many systems [6]. There are many ways to store and organize data. The hardware level has the storage mode set for the speed difference between the cache and the hard disk, and the software level stores the data separately according to the reading and writing, so as to reduce the pressure of the database. In fact, the development of more common is the sub table and partition storage technology. Sub table and partition are similar, the purpose is to improve the operation efficiency, both are very similar, both need rule decomposition table [7]. The difference is that table splitting divides a large table into multiple entity tables, while partition is for data segments. There are many ways to evaluate the data. Generally speaking, it is mainly in the following four aspects: the consistency of the data, whether the data information is accurate enough, whether the data is complete and whether the data is excessive, and the timeliness is poor [8]. Based on this, the distributed data and traditional database are analyzed and designed. Distributed storage will combine the data characteristics of blockchain and select mature NoSQL database as the basic storage. Blockchain is designed for massive data registration, including static data and dynamic data. Static data is the information registered in the form of metadata to manage the registration data. Facing the storage bottleneck of massive data environment, both static data and dynamic data must adopt distributed storage [9]. The data partition repository structure of the network center is as follows (Fig. 1):

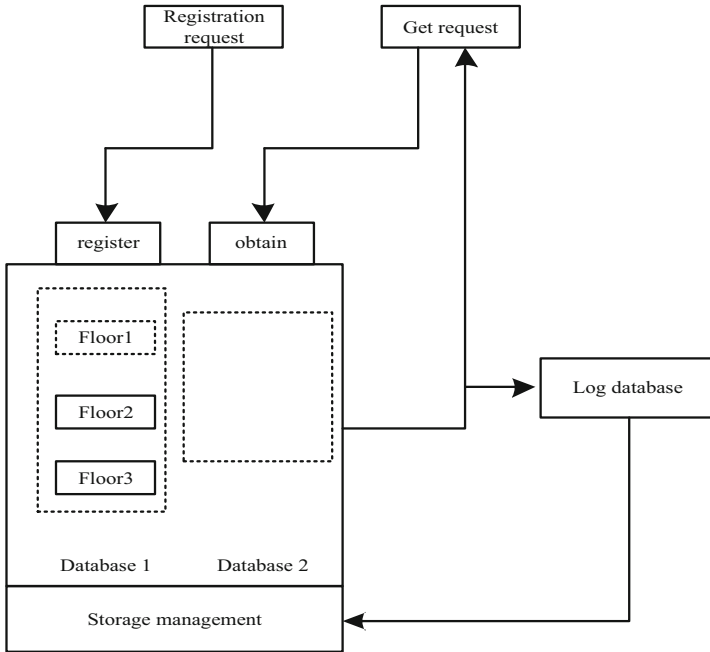


Fig. 1. Data partition repository structure of Network Center

Data consistency algorithm for different scenarios, the distributed consistency of mass data is mainly the consistency of multiple copies. The current methods are also relatively diverse. The consistency methods for distributed systems mainly have solutions, mainly with the help of transaction control method to achieve the effect, as well as the extended transaction control method based on the extension, through the data replication The replication control method for controlling, the message queue method for controlling messages, and the primary replica update message queue method. It is self-evident that the accuracy of data is only the basic standard of information availability, and the distorted data itself can be regarded as the data without quality. Most of the researches in this field are in the state of model and lack of uniform applicability. From the perspective of decision maker, the timeliness of data is related to the accuracy of decision. From the perspective of ordinary users, timeliness is related to the selectivity of users. From the perspective of system, timeliness is related to the offset of system overhead. Complete the data balancing partition, and its basic composition is shown in the figure below (Fig. 2).

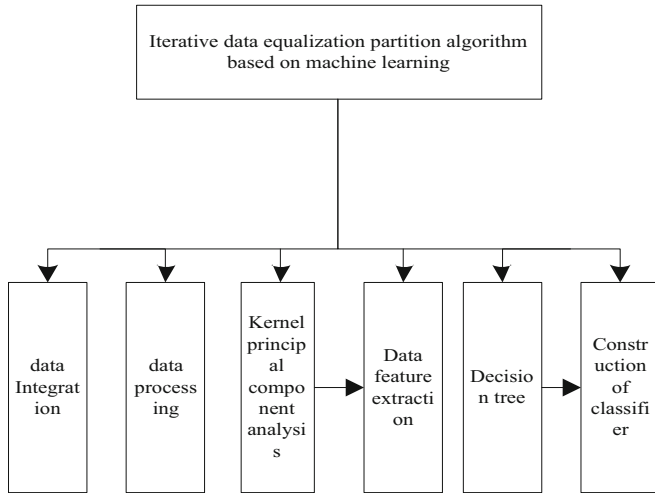


Fig. 2. Composition of data balanced partition structure

It can be seen from the figure that the iterative data equalization partition algorithm based on machine learning includes three parts.

In the first part, heterogeneous data is integrated and processed to eliminate invalid data and reduce data scale.

The second part is to extract data features.

In the third part, the classifier model is constructed to realize the data Iterative balanced partition.

As a logical data pool, blockchain is the storage place for data registration and storage. As shown in the figure, as the core of data oriented architecture, blockchain is responsible for internal data management and external interaction to provide access and location services. “Physical” cloud refers to the data cloud of the physical world, such

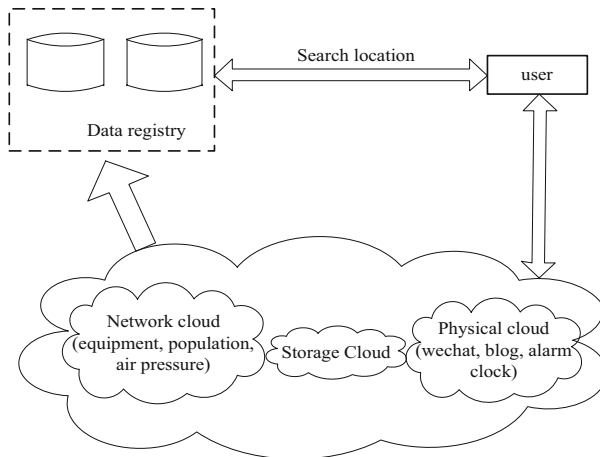
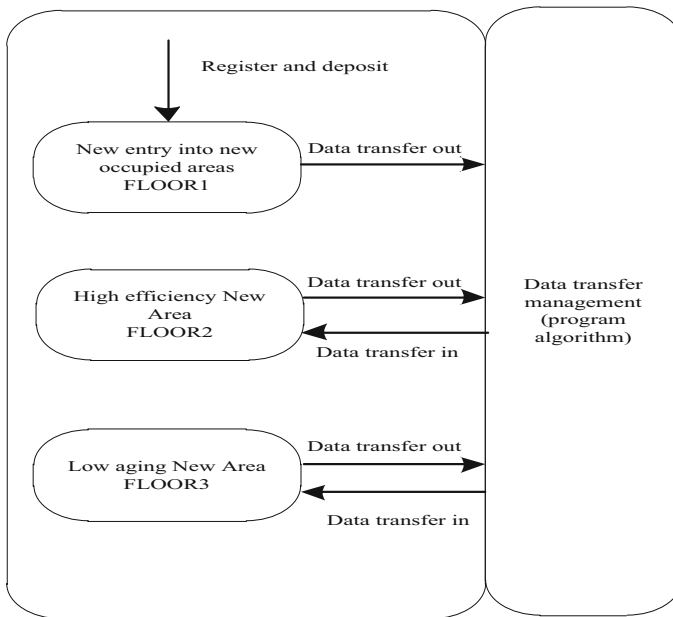


Fig. 3. Data partition processing logic of Network Center

as all kinds of natural substances and all kinds of entity attributes; “network” cloud refers to all kinds of attribute data cloud in the network (Fig. 3).

It further shows that the dynamic data forms a chain structure according to the time. Here, the static data has two operation records of “dynamic data 1” and “dynamic data 2”, which are arranged in turn. It can also be seen that the dynamic data tracking is the static data information, and the design is to present the chain structure in turn according to time. The unique identifier is used as the link relationship identifier. With the help of dynamic data, we can fully obtain the change process of static registration information, and track the operation change of data according to time. This will play an auxiliary role in the authority supervision of the later authority components. The relationship between dynamic data is shown in the Fig. 4.



**Fig. 4.** Dynamic data relationship

In view of the rapid growth of data, as a distributed database, blockchain adopts automatic fragmentation mechanism to realize the distributed storage of massive data. In this paper, we will analyze the fragmentation mechanism of the blockchain, and design the patch key based on the characteristics of the blockchain itself and the wide range of data access statistics. At the end of data integration and processing, and as a sample data, data feature extraction. Based on principal component analysis, the kernel method is applied to form the kernel principal component analysis method. A set of data to be analyzed is mapped into a suitable high-dimensional feature space according to a nonlinear mapping rule, and then the data is processed and analyzed in this space by using a linear learner



$$O(x_i, y_i) = \sum_{n=1}^k \zeta(x_i, y_i)n \tag{1}$$

In the formula,  $O(x_i, y_i)$  is the coordinate of the data in the high-dimensional feature space after nonlinear mapping;  $k$  is the space dimension;  $n$  is the number of data in the data sample;  $\zeta$  is the nonlinear mapping rule, that is, the mapping function.

Assuming that the order of the curve filled in the massive data space is  $M$ , the  $S$  space of the massive data set can be divided into  $2^M \times 2^M$  grids, and each grid has a four-dimensional spatial Hilbert code:

$$M_0 = O(x_i, y_i)H \left[ \log_2 \frac{D_0}{H_1} \right] 2^M \times 2^M \tag{2}$$

In the formula:  $D_0$  represents the total amount of data;  $H_1$  represents the storage size of data block. Statistical coding data element information set  $O$ , assuming that the total number of data coding blocks is  $i$ , if the data block storage size  $H$  is greater than the maximum threshold percentage of massive data blocks, then the coding fast is divided into sample set  $y_i$ .

According to the aggregation characteristics of spatial four-dimensional Hilbert coding linear filling curve, the massive data coding blocks are decomposed, and the corresponding storage sequence of each coding block  $H_2$  is marked to form the corresponding spatial data partition matrix, as follows:

$$F = \begin{vmatrix} H_{2de0} & H_{1a0} & s_0 \\ \dots & \dots & \dots \\ H_{2den} & H_{1an} & s_n \end{vmatrix} \tag{3}$$

In the formula:  $H_{2den}, H_{1an}, s_n$  represents the spatial element of massive data, according to which the corresponding spatial code and the corresponding massive data block storage label after matrix matching are obtained, thus the spatial element division of massive data is completed.

According to one of the items, the corresponding massive data set is divided. Assuming that the length is larger than the width, then the horizontal axis direction of massive data is the positive direction set, and the vertical axis direction is the opposite direction set, so the spatial element interval of massive data is calculated [10]. Through this interval, the coding block is decomposed to realize the feature partition of massive data. Combined with the organization layer technology of relational database, the three-stage aging model is designed. The design focuses on the reasonable distribution of data, the design of data aging algorithm, and the three-tier design of storage area according to the aging [11]. In the face of large amount of data and high throughput conditions, there are two basic methods to solve the problem of insufficient compression performance: vertical expansion and fragmentation. Monodb adopts fragmentation mechanism.

The fragmentation mechanism of blockchain is shown in the figure data fragmentation [12]. Data is divided into multiple places to share the storage, which is the

basis of massive data storage in the system. There are mainly two ways of slicing: by range and by hash (Fig. 5).

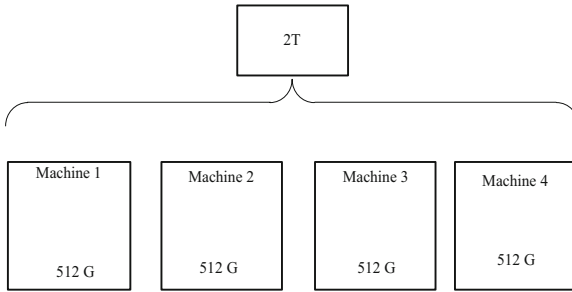


Fig. 5. Data fragmentation processing method

Storage control is responsible for data transfer among them. The three layers are new data area, high aging data area and low aging data area.

### 2.3 Implementation of Data Partition Storage

The data source is not necessarily uniform, and the format is not necessarily the same, which is not conducive to partitioning and reduces the speed of partitioning, so it is necessary to integrate these heterogeneous data before partitioning. Data integration is the process of integrating data of different sources, formats and characteristics into one big data, so as to serve the subsequent partition. Here, data integration is mainly completed by central data warehouse technology, and its structure is shown in the figure below (Fig. 6).

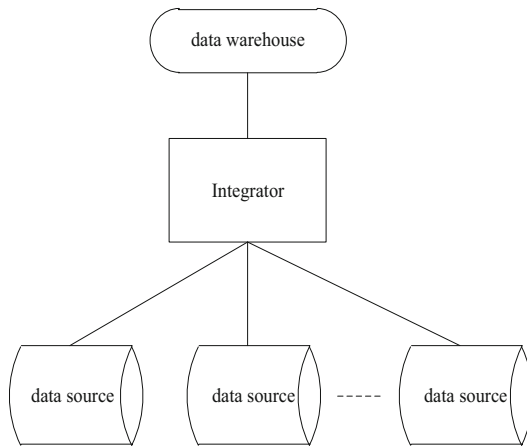


Fig. 6. Data partition processing library

The biggest advantage of central data warehouse is that it can achieve the maximum control of the extracted data, and its application effectively solves the problems of data dispersion, diversity and redundancy.

After data integration, further data processing is needed, including data cleaning and data reduction.

Data cleaning is to find and correct abnormal data in data set, including data consistency check, invalid value and missing value processing, duplicate data deletion, etc.

Data reduction refers to reducing the amount of data and reducing the scale of data. The amount of data after integration is huge. If it is directly used in the follow-up work, it will not only increase the amount of data, but also reduce the accuracy of partition. Therefore, it is necessary to reduce the original data after integration, including data aggregation, dimension reduction, data compression, data block reduction, etc.

The timeliness of data is an important aspect of data value. In the past, the judgment of data timeliness was mainly based on the fact that the data was only valid in a certain period of time. But for the generalized data, it is not enough to only consider this point. There are not many algorithms to determine the timeliness of data, but the design is to consider the efficiency of data query and the cost of system resources. The three-stage partition setting of the system has different overhead configurations for different partitions. In order to improve the adaptability and controllability of the system, the following time effective algorithm is designed on the basis of the original “one shot dead” and “dominated by the user”, so as to achieve the function of controllable allocation of index resources. In the design of timeliness judgment algorithm, the controllability of timeliness is mainly considered.

$$N = M_0i + 3TF \tag{4}$$

$$f(x) = a_0 + \sum_{N=i} (f_1(c_i, c_{i+1}) * f_2(c_i, c_{i+1})) \tag{5}$$

$$f_1(c_i, c_{i+1}) = \text{sgn}(c_{i+1} - c_i) \tag{6}$$

$$f_2(c_i, c_{i+1}) = \frac{|c_i - c_{i+1}|}{\max(i, c_{i+1})} \tag{7}$$

In the formula:  $a_0$  is the reference value, as the floating reference.  $c_i$  here is the number of visits in the T cycle.

The number of visits in four consecutive t-cycle time in turn. T here for a day or more. The longer it is, the easier it is to cover the downlink line, and the slower it is to respond to the short-term steep drop in the number of visits. The segmentation value of high and low timeliness is L. The aging value above L is high aging, and below L is low aging. Set the high efficiency value as a. Above a value is high aging, below a value is low aging.

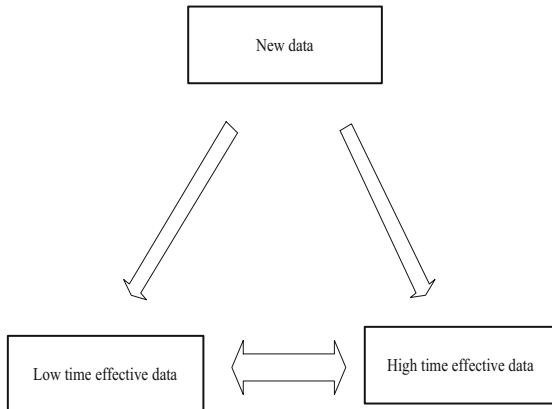
$$\begin{cases} \text{sgn}(f(x) - A) = 1 \\ \text{sgn}(f(x) - A) = -1 \end{cases} \quad (8)$$

For the first new data area, there is a regular transfer out, the amount of data is relatively stable, there will be no data accumulation. This area is indexed on keywords according to the query. The source of data in this area is registration. Transfer out includes two directions, one is the second layer of efficient data area, the other is the third layer of inefficient data area. The direction of transfer out depends on the storage management module. According to a judgment calculation result of aging algorithm, if the timeliness of T cycle is good, it will enter the high aging region, otherwise it will enter the low aging region. For the efficient data area of the second area, the data of this area is also in and out, and the volume is relatively stable. The worst case is that all data is often used, resulting in data accumulation here. According to the data characteristics of the three regions, the data resources are allocated. Index is very important for data query performance, and it costs a lot. When the index size exceeds the memory capacity of the system, it will bring too much V0 performance consumption. So there is a tendency in the configuration (Table 3).

**Table 3.** Resource allocation

Hierarchy	Configuration policy
FLOOR1	Configure keyword index, title index and subject
FLOOR2	Configure keyword index, title index and subject
FLOOR3	Configure title index

Further optimize the data partition storage state, as shown in the Fig. 7:



**Fig. 7.** Data partition storage status

The status of data can be divided into new data, high aging and low aging. The area the data should belong to is determined by its aging value. Different from the other two areas, new data can only be registered in one data entry mode. According to the access situation in a certain period of time, there are two destinations. The data of the two outer regions flow through each other, and the data from the new data area can be obtained. The timeliness of data is an important aspect of data value. In the past, the judgment of data timeliness was mainly based on the fact that the data was only valid in a certain period of time. But for the generalized data, it is not enough to only consider this point. Not all the data.

### 3 Analysis of Experimental Results

In order to test the effectiveness of data balancing partition algorithm, simulation test is needed. First of all, the test needs to clarify the experimental environment, as shown in the following Table 4.

**Table 4.** Simulation test environment

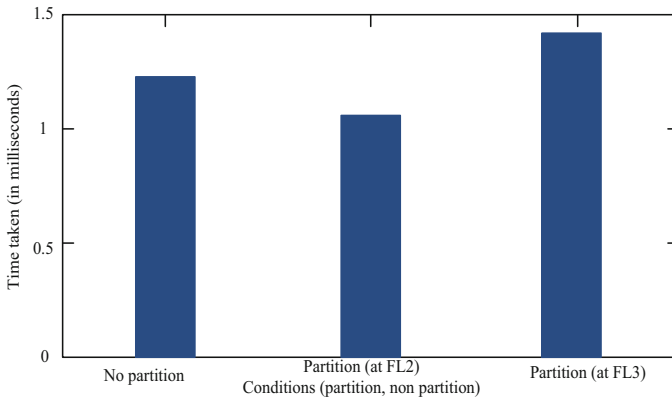
Name	Parameter setting
Operating system	Windows 764 bit
CPU	Intel, corei5-24110m memory 4G
Application server	Ggda-25t5 special server
Server database	Oracle 10g
Simulation experiment software	Matlab R2009a
Test recording tool	Dev test tool

The storage balance test scheme is shown in the Table 5.

**Table 5.** Storage balance test scheme

Test items	Storage balance test
Test purpose	Test whether the data put in is balanced
Prefabrication conditions for testing	1. Subject and_ ID is the partition key
	2. _ ID system insertion. Subject randomly generates character strings
	3. Each block is limited to 1m
	4. Set two servers as two partitions
Testing procedure	1. Take FL2 as the object, insert 10000000 pieces of metadata
	2. Check the blocking condition
Expected results	The data block distribution of the two partitions should be relatively balanced

In order to test the data equalization ability, the segmentation limit of each data block is 1m. Insert 10 million pieces of data, data storage balance display shard0000, shard0001 each contains 796 and 797 data blocks. In the experiment, the size of each block is limited to 1m, and the amount of data stored in each chip is basically the same. The test scheme is shown in the table. The time-consuming statistics of the measured results of the network center data partition test scheme are shown in the Fig. 8.



**Fig. 8.** Comparative analysis of partition storage time consumption

The basis of data partition in network center is that the active data is a small part, which can save time when querying a small range. It is proved that the block chain is more effective in the network center data partition storage model.

## 4 Conclusion

Under the background of big data era, the development of network information technology makes it possible that all kinds of information is no longer an information island. In view of the rapid growth of data, how to form a data oriented development system with data as the core is of great significance to data sharing and management. Based on this, combined with the blockchain to build the network center data partition storage model, in order to better deal with massive data effectively. Although the mathematical model constructed in this paper can effectively deal with network center data, there are still many shortcomings. In the future, we will study the changing attributes of network center data and the increasing amount of data to improve the performance of this method.

## References

1. Li, X., Jiang, P., Chen, T., et al.: A survey on the security of blockchain systems - ScienceDirect. *Future Gener. Comput. Syst.***107**(4), 841–853 (2020)

2. Fraga-Lamas, P., Fernandez-Carames, T.M.: Fake News, disinformation, and deepfakes: leveraging distributed ledger technologies and blockchain to combat digital deception and counterfeit reality. *IT Prof.* **22**(2), 53–59 (2020)
3. Gao, X., Ren, B., Zhang, H., et al.: An ensemble imbalanced classification method based on model dynamic selection driven by data partition hybrid sampling. *Expert Syst. Appl.* **160** (5), 113–115 (2020)
4. Nomura, K., Shiraishi, Y., Mohri, M., et al.: Secure association rule mining on vertically partitioned data using private-set intersection. *IEEE Access* **1** (2020)
5. Levitin, G., Xing, L., Huang, H.Z.: Security of separated data in cloud systems with competing attack detection and data theft processes. *Risk Anal.* **39**(4), 846–858 (2019)
6. Nozal, R., Perez, B., Luis Bosque, J., et al.: Load balancing in a heterogeneous world: CPU-Xeon Phi co-execution of data-parallel kernels. *J. Supercomput.* **75**(3), 1123–1136 (2019)
7. Nguyen, T.D., Lee, S.W.: PB-NVM: a high performance partitioned buffer on NVDIMM. *J. Syst. Arch.* **97**(11), 20–33 (2019)
8. Shanthi, P.A., et al.: Privacy preserving time efficient access control aware keyword search over encrypted data on cloud storage. *Wirel. Pers. Commun.* **109**(4), 2133–2145 (2019)
9. Zhang, Y.F., Wang, X.P., Pan, Y.X., et al.: Alteration in isotopic composition of gross rainfall as it is being partitioned into throughfall and stemflow by xerophytic shrub canopies within water-limited arid desert ecosystems. *Sci. Total Environ.* **692**(20), 631–639 (2019)
10. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Compl. Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
11. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. *Front. Genet.* **10**, 120 (2019)
12. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. *Entropy* **21**(9), 902–908 (2019)



# Enterprise Cluster Intelligent Manufacturing Information Management System Based on Wireless Communication Technology

Yang Su<sup>1</sup>, Yu-hua Gai<sup>1</sup>, and Qiu-ying Lv<sup>2</sup>(✉)

<sup>1</sup> School of Economics and Management,  
Yantai Institute of Technology, Yantai 264005, China

<sup>2</sup> School of Business, Fuyang Normal University, Fuyang 236041, China  
cr.njb1121@alipay.com

**Abstract.** The existing intelligent manufacturing information management system of enterprise cluster has not reached the optimal path when designing the system security. In order to improve the security of the system, the intelligent manufacturing information management system of enterprise cluster is designed based on wireless communication technology. In the design of hardware system, the information processing module is designed to reduce the time used for data reading and writing, and the circuit system of wireless communication technology is designed to enhance the security of data transmission. In the design of software system, the whole function of the system is designed, the optimal path algorithm is designed, and the security of information management algorithm is strengthened. Through experiments, we can see that the function of the intelligent manufacturing information management system of the enterprise cluster is tested, and the integrity of product data management module, system functional framework module and production implementation management module is tested. The security of the system is tested and it is found that the system has enough security when the data transmission unit is below 25 bits.

**Keyword:** Wireless communication technology · Enterprise cluster · Intelligent manufacturing · Information management system

## 1 Introduction

Enterprise cluster usually refers to a large enterprise with many related industries, or many content related industries are associated with each other in the form of a collection. This combined operation mode can make the enterprise have better work effect. In actual production, due to the increase of production scale, in order to have enough work efficiency, enterprises usually use intelligent technology to optimize the manufacturing process. Technicians can use network technology and automation technology to complete the management of intelligent manufacturing related information, and use the efficient collection of production information to improve the quality of management work [1]. In the process of enterprise development, due to the influence of various external technologies, it needs to keep pace with the times in the development. Therefore, it is necessary to design a fast and convenient enterprise cluster intelligent



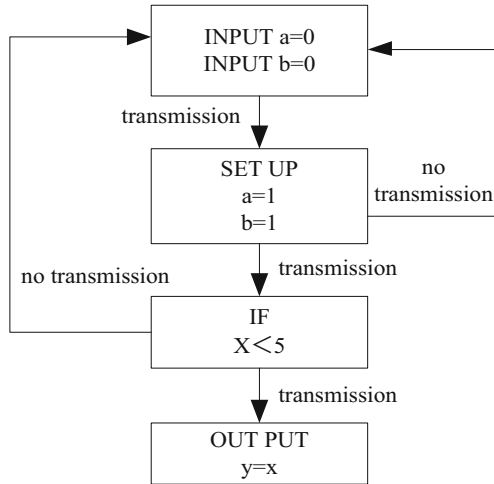
manufacturing information management system, through which the enterprise related information can be counted and managed to realize the rapid development of the enterprise. At present, the common systems include intelligent manufacturing system based on MES. Aiming at the problem that the low level of manufacturing process informatization restricts the production efficiency of small and medium-sized enterprises, this paper designs and develops an intelligent manufacturing system based on MES. The system completes the rapid data exchange and process control through industrial Ethernet, uses MES to collect the operation information and working status of all equipment, real-time deployment of the process, and realizes the storage and reclaiming of the system, Intelligent control of production process such as detection and identification, sorting and positioning; So as to realize the informatization and digitization of manufacturing process, and effectively improve the operation efficiency of enterprises. Advanced intelligent manufacturing plan management system based on digital twin. The plan management system can't adjust automatically according to the uncertain factors under the mode of multi variety and small batch manufacturing, A framework of intelligent manufacturing planning management system based on digital twin is proposed. The PDCA business process model of planning management is established. The processing logic and constraints of the cycle rolling model of static scheduling and the optimization model of dynamic emergency scheduling are proposed, It is verified that the management mode proposed in this paper can support the dynamic adjustment of intelligent plans with different orders and resource status, and improve the efficiency of dealing with uncertain factors. Although these two more advanced systems effectively promote the development of enterprises, there are problems of poor communication quality and poor real-time update speed in the process of using them.

Therefore, in the current development of manufacturing industry, the research of enterprise cluster intelligent manufacturing information management system plays an important role. Most of the existing information management systems only focus on the management function and operation implementation action requirements, but ignore the security requirements of the system itself. Therefore, this paper designs the enterprise cluster intelligent manufacturing information management system based on wireless communication technology, and makes a new design for its security.

## **2 Design of Enterprise Cluster Intelligent Manufacturing Information Management System Hardware Based on Wireless Communication Technology**

### **2.1 Design Information Processing Module**

In the hardware design of enterprise cluster intelligent manufacturing information management system, the response mechanism of information fast processing module is the most core peripheral module, and its composition structure is shown in Fig. 1.

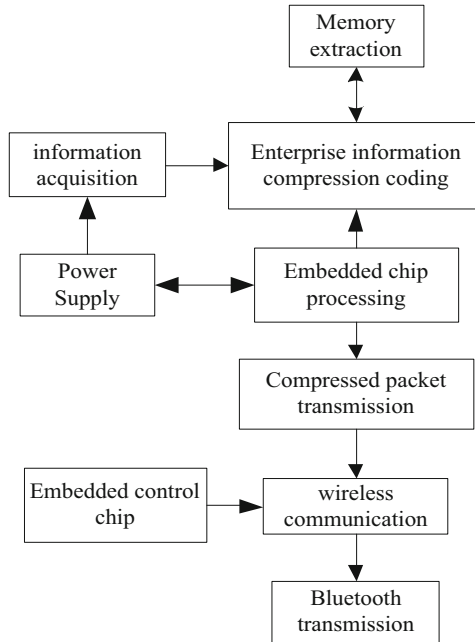


**Fig. 1.** System hardware fast response conversion module

As shown in Fig. 1, in the bus with equal bandwidth, if the ABP bridge in neutral position is used as the main template, if the three states of space, setup and output can be perfectly controlled, the read and write data can be quickly converted to the safe area and transmitted to the designated location [2]. In the timing chart of read-write transmission, the address information is consistent with the selection signal. When the status of the selection prompt box reaches a certain degree, it means that the read-write transmission data can be up sampled, and the reading sampling time will be improved to a certain extent compared with the traditional sampling storage time.

## 2.2 Circuit Design of Wireless Communication Technology

In the design of wireless communication module, the most important is to collect information based on wireless communication technology. The main control center of the system is a kind of control chip which can realize information collection, and also has certain operation transplant function. Embedding such chip into the wireless communication information collection system can greatly enhance the management ability of the operating system. In the embedded chip setting, the power management ability of the visual acquisition system and the design of low power consumption management program can be realized through low power energy conversion. The overall framework of the information acquisition system is shown in Fig. 2.



**Fig. 2.** Overall design of wireless communication technology

As shown in Fig. 2, the embedded chip management system extracts memory and transmits compressed packets through a compressed and coded information management mode. When the transmission efficiency of enterprise information is more than 50 FPS (Frames Per Second), the energy loss of image transmission can be converted. In this process, if the information management system can complete the analog conversion of clock generator or A/D (Analog.Digital) conversion, the output mode of information management signal can be controlled. Through Bluetooth transmission interface, the resolution setting of video program compression and coding can adjust the process of information transmission, that is, it can be equipped with a 220 V power supply in the chip of embedded information acquisition system. In the control end, the communication connection of Bluetooth transmission function is also an important step. This information transmission program based on embedded chip can provide a compatibility channel for the existing Bluetooth devices, and its YSB interface can realize the physical connection with the host of the computer device control end, and then realize the connection channel between voice and text conversion.

In the initial circuit design, the above embedded chip information acquisition system provides part of the power output conditions. At this time, the output device adopts 3.0 V standard voltage.

However, after a certain period of testing, it is found that the overall design of the circuit as shown in Fig. 3 needs to achieve the purpose of grounding anti-interference through the zero point group, so different power line numbers need to be used to make these grounding lines bypass the interference and complete the information transmission

and identification of the whole system module [3]. In this process, two 30  $\Omega$  resistors can be used as the matching resistor of Bluetooth transmission module, and five external grounding power supplies can be used as the output of line impedance.

### 3 Design Enterprise Cluster Intelligent Manufacturing Information Management System Software

#### 3.1 Overall Function Design of the System

First, in the development and design of manufacturing data management system, technicians should first complete the perfect database design to meet the needs of subsequent production management. In the production management, the integrated information management system needs to compile, modify, delete and retrieve the data table combined with the collected data [4]. In the actual design and research, the above functions can use SQL (Structured Query Language) database to complete the collection and storage of production attribute data, and Mongo DB (Data Base) to complete the storage of production implementation data [5]. In the attribute database, the main data content includes production raw material data, product data, order data, order product data, warehouse data, equipment operation data and function staff data. The following table is the main data list in database design.

Second, the design of data receiving module. Data receiving module can use PDM as the basis of design in product data management. If the manager needs to change the product data, the management system can send the change data to the PT system, and then the production management department can complete the change of the raw materials and parts data in production combined with the data to be changed, so as to complete the data collection and change.

Third, data acquisition module. The technical basis of data acquisition module is OPC (OLE for Process Control) technology, which is connected with manufacturing equipment by means of multi-layer interface, and completes the exchange of production data by means of server and network to obtain real-time data of manufacturing process. In the current design, production data is usually collected by periodic sampling and subscription [6]. Periodic sampling data is information collection, the client obtains the data of the server within a certain time and frequency, and the subscription method is to automatically send the data changes with the help of the server, which reduces the workload of the server. In data acquisition, different types of information can be collected in different ways. For the signal data with fixed change period such as temperature and speed, technicians can use the periodic sampling method. For some sudden information such as alarm data, technicians generally choose the subscription mode to improve the real-time of on-site data acquisition.

Fourth, data transmission module. The main function of the data transmission module is to push and transmit the data of the production site, so as to realize the real-time supervision of the production site. The data acquisition module will exchange data with the help of network technology and server, so as to complete the real-time push of client information and ensure the normal management work [7]. The functions of data transmission module include creating information transmission channel, real-time data transmission and channel automatic closing, which ensure the normal management.

Fifth, the general assembly business module. The first mock exam of the module includes the editing of the parts used in the production process, the management of the equipment commands and the production of raw materials in the manufacturing process, and the data collection and management of material consumption and production cost in the manufacturing process, so as to ensure the correct understanding of the whole production process and ensure the high quality of the management work.

Sixth, sales management module. The integrated information management system also needs to manage the sales process of the products. The system can use ASP platform to complete the processing of order data. After the order information analysis is completed, the marketing department can configure the products according to the demand of the order. The configuration process shall be carried out strictly according to the customer order requirements to ensure the marketing quality [8]. After the product is configured, the system can form the only product production code. The marketing department can use this code to statistics and manage the orders, and form the demand plan by integrating the data to ensure the comprehensive management quality of the system.

Seventh, purchasing management module. The main purpose of this function module is to combine with other system modules to complete the whole process management of material transportation and use. The procurement management module is mainly used to collect and calculate the data of raw material procurement, supply and transportation, which helps to strengthen the management process of materials, reduce unnecessary production costs, and improve the manufacturing and production benefits of enterprises.

Eighth, data comprehensive query module. In the information management process of manufacturing process, the system should not only complete the collection, transmission and processing of all kinds of manufacturing information, but also have the function of retrieval and query, which is convenient for managers to query some previous data and optimize the manufacturing process [9]. The retrieval function of the system ensures that users can search the corresponding information, and complete the efficient management of the whole process of manufacturing with the help of a unified data system. In the design of the system, in order to ensure the safety of management, enterprises can set certain operation authority to avoid some criminals stealing production information.

### 3.2 Security Design of Information Management Algorithm

Its principle A\* search algorithm, when analyzing and introducing the algorithm, the heuristic search algorithm should be introduced first. The algorithm can search for other nodes based on known nodes, and can obtain other related methods based on heuristic functions, calculate and select the safest node as the next search node. When searching for other nodes, both bfs (Basic Frame Synchronization) and dfs (Distributed file system) use a blind search method, so they will not select nodes through heuristic functions. If you are unlucky, you may need to test all the solution set spaces before you can get the results. Therefore, this algorithm is generally only used in clear conditions where there is only a small solution space. There are obvious differences between Bfs, dfs and heuristic search. The heuristic search algorithm is to obtain the

optimal solution or better solution in a very short time. A solution can be obtained by calculating the heuristic function. How to design this heuristic function is the key. The  $A^*$  ( $A - Star$ ) search algorithm is mainly to find the safest path in a graph with multiple nodes. It is a heuristic search algorithm. The formula is:

$$f(n) = g(n) + h(n) \tag{1}$$

Among them,  $h(n)$ ,  $g(n)$  and  $f(n)$  respectively represent the estimated cost, the actual cost and the function from the target node to the departure node. The  $A^*$  algorithm will use a structure similar to the bfs algorithm to select the priority queue during calculation, and determine the value of the fns of the possible child nodes by calculation, and will stop the calculation until the target child node or the priority queue is determined [10]. If the estimated cost value of the  $A^*$  algorithm is 0, it is relatively similar to the DFS algorithm, and if the actual cost value of the  $A^*$  algorithm is 0, then it is relatively similar to the bfs algorithm. In the former case, only calculation of  $g(n)$  is needed to determine the safest path. The most distinctive feature of the  $A^*$  algorithm lies in the design of its evaluation function. Therefore, the quality of the  $h(n)$  design directly affects the efficiency of the  $A^*$  algorithm.

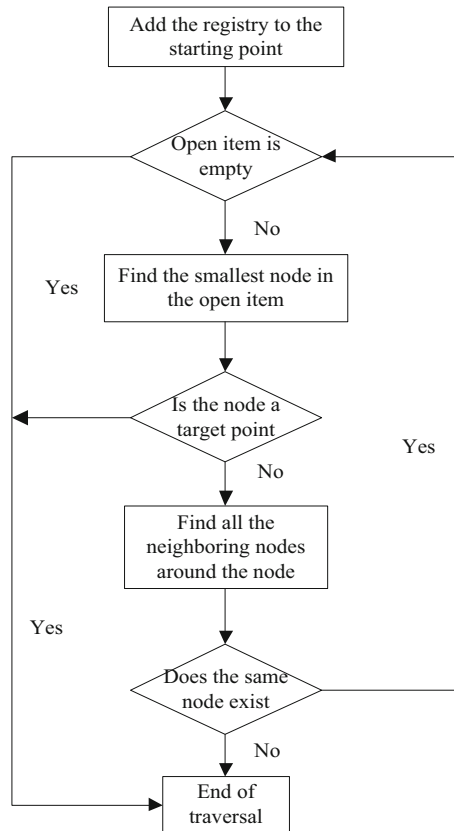


Fig. 3. Algorithm design

As shown in Fig. 3, the principle of the  $A^*$  algorithm is that when selecting subsequent traversed nodes, it is usually completed by each node that has not been traversed, but the node with the least value is selected to replace the node with the smallest value among the detected nodes. For example: Find the shortcut of  $V_0 \rightarrow V_5$ . In the process of  $V_0 \rightarrow V_5$ , you can reach the final node  $V_5$  through  $V_1$ ,  $V_2$ ,  $V_3$ , and  $V_4$ . When solving the best path in a static network, the  $A^*$  algorithm is a direct and effective algorithm. Whether the valuation function can be selected reasonably will directly affect whether the shortest path can be accurately determined. The similarity between the actual value and the estimated value can reflect the selection of the valuation function. If the value of the evaluation function is 0, then only the calculation for  $g(n)$  is needed to determine the minimum path. However, the smaller the evaluation function, the more nodes, which will increase the workload of calculation and consume more time in calculations, that is, the smaller the evaluation function, the higher the security of the algorithm.

## 4 Experimental Study

### 4.1 Experiment Preparation

This experiment compares the system design in the article with the traditional enterprise cluster intelligent manufacturing information management system to determine whether the system designed in this article is practical and safe. The environment configuration used in the experiment is as follows: CPU uses Intel(R) i5-2440@3.40 GHz processor, running memory is 4 GB, and storage space is 128G. The operating system of its software platform is Windows 10 64 bit, the web database is established through Tomcat 6.0 software, its back-end database uses MySQL, and the MyEclipse system is used for back-end development in the J2EE environment, and the back-end program development language is Java.

The design of the online teaching system for financial management courses in this article is based on the B/S structure, mainly the information interaction between the browser, the WEB server and the database. The user sends a request to the WEB server through the client browser, and the WEB server receives the request and It exchanges information with the database. After receiving the request from the WEB server, the database performs corresponding operations and returns the running result to the WEB server. The WEB server generates code and returns it to the client browser. The browser interprets the HTML code and displays the content of the web page. After setting up the above experimental environment, you can start the test of system functions and system performance.

### 4.2 System Function Test

Functional testing is the most basic test in system testing. It does not matter how the software is implemented internally, but only verifies whether the product's functions meet the requirements specifications based on the requirements specification and the test requirements list. It mainly checks the following aspects: First, function Whether it

is fully realized, and whether there are any omissions; second, whether the function meets the hidden requirements of user needs and system design; third, whether the input can be correctly accepted and the correct result is given. The specific test situation is shown in Table 1.

**Table 1.** Functional test results

System module	Subsystem	Test results
Product data management module	Operating system management	Normal
	Product structure management	Normal
	Product configuration management	Normal
	Material management	Normal
	System maintenance management	Normal
	Manufacturing process management	Normal
	Product code management	Normal
System function framework module	Personal work interface	Normal
	Information management system	Normal
	Project management system	Normal
	Process management system	Normal
Production implementation management module	Real-time data collection system	Normal
	Data management system	Normal
	Data display system	Normal

In summary, the test results show that each function can be used normally, and the function is complete, the interface is friendly, and the interaction is good.

### 4.3 System Security Test

The enterprise cluster intelligent manufacturing information management system, as a B/S-structured Web system, not only needs to have beautiful pages, practical functions, but also strong security requirements, otherwise it will cause data loss and system paralysis. Because this system is directly facing Internet users, it may be threatened by more attacks on the network, and it is very necessary to test its security. The goal of security testing is to find the security flaws in the system design as much as possible, and then make the necessary repairs. The security test of this system includes testing whether there are security vulnerabilities in the login of the system, testing whether



there are security vulnerabilities in the script program on the server side, testing whether the system management code is safe, preventing users from bypassing the program and directly accessing the background program, testing the Tomcat system server platform Whether the security settings are appropriate and so on.

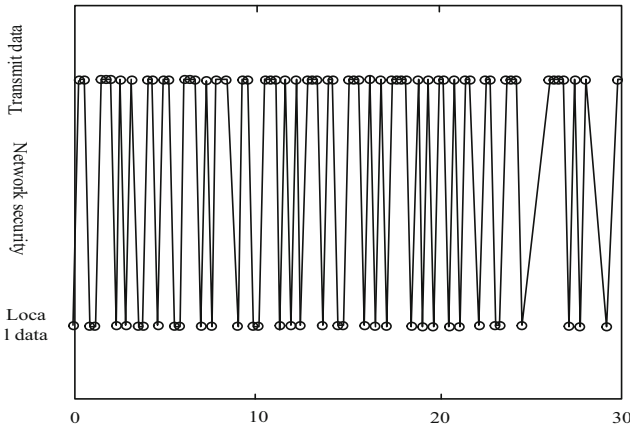
Take security as the evaluation index of the enterprise cluster intelligent manufacturing information management system, as the basis for analyzing system performance, perform system classification performance testing, and analyze the integration capabilities of different systems for information resources. The formula for calculating the security of the enterprise cluster intelligent manufacturing information management system based on wireless communication technology is as follows:

$$Q_i = \sum_t^{N_i} b_t(i)N_i \quad (2)$$

In the formula,  $Q_i$  represents the probability of the information management system being invaded by malicious data;  $N_i$  represents the number of samples in the data set;  $b_t(i)$  represents the total number of samples in the  $i$  data currently running. To extract the characteristic quantities of the security performance of the enterprise cluster intelligent manufacturing information management system, the following calculations are required:

$$G_{m,n} = \begin{bmatrix} C_r^b(q_{k+1}) & 0 \\ 0 & C_r^b(q_{k+1}) \end{bmatrix} \begin{bmatrix} r_a \\ r_m \end{bmatrix} + \begin{bmatrix} v_{k+1}^a \\ v_{k+1}^m \end{bmatrix} \quad (3)$$

In the formula,  $C_r^b(q_{k+1})$  represents the serial probability value of the enterprise cluster intelligent manufacturing information management system based on wireless communication technology being illegally invaded;  $q_{k+1}$  represents the matching coefficient of the security performance of the enterprise cluster intelligent manufacturing information management system; in the above formula, it will usually be running The module and the influence of non-networked data are divided into 2\*2 sections, and  $v_{k+1}^a$  and  $v_{k+1}^m$  are expressed as the coefficient components of the dynamic characteristics of the network. According to the above calculation formula, it is possible to obtain a schematic diagram of the evaluation results of the security performance of the wireless connection terminal data and local data during the process of networking in the enterprise cluster intelligent manufacturing information management system, as the number of data runs increases.



**Fig. 4.** Data security performance evaluation results

As shown in Fig. 4, with the increase in the number of data operations, the network security of the enterprise cluster intelligent manufacturing information management system fluctuates to a certain extent. Compared with the data at the wireless transmission end, the local data is safer. But in general, before the number of data runs does not exceed 25 groups, the risk is still within the range that can be borne. Therefore, the same user needs to pay attention to the simultaneous use of data not exceeding 25 when using the enterprise cluster intelligent manufacturing information management system. Group, otherwise it is easy to cause network security accidents.

## 5 Conclusion

In the above, an enterprise cluster intelligent manufacturing information management system based on wireless communication technology is designed, and the function test of the system in the experimental link proves that the software design has complete functions, friendly man-machine interface, and simple use. The operation process is safe and reliable, the use effect is good, and it has good cross-platform security, and basically achieves the expected goal. However, due to limited time and energy, some of the design functions of this system have not been perfected. This will continue to be improved in the later stage. In addition, the interface needs to be further beautified to meet the needs of employees in different departments.

**Fund Projects.** “This article is the research result of the special project of Shandong Province Social Science Planning “Intelligent Manufacturing Promotes the Integration of Internet, Artificial Intelligence and the Real Economy” (No. 20CSTJ24).

## References

1. Zhang, L., Li, M., Wei, S.: Development of information management system in the construction of digital forest farm. *Manag. Technol. SME* **3**, 28–29 (2021)
2. Liu, G., Wang, J., Li, Y., et al.: “One graph of power grid” spatio-temporal information management System. *Electr. Power Inf. Commun. Technol.* **18**(1), 7–17 (2020)
3. Zhang, Y., Zhu, Y., Wu, M., et al.: Simulation of complex network model for innovation diffusion dynamics in intelligent manufacturing cluster in small and medium-sized enterprises. *J. Qingdao Univ. Sci. Technol. (Soc. Sci.)* **35**(1), 35–40 (2019)
4. Liu, G., Chen, M., Liu, S.: Thoughts on the cultivation and development of precision instrument and equipment industrial cluster in Guangdong. *Mech. Electr. Eng. Technol.* **49**(6), 1–3 (2020)
5. Xiang, J., Tang, J.: Construction of course system for intelligent manufacturing-oriented multi-axis CNC machining specialty in secondary vocational schools. *Jiaoyu Jiaoxue Luntan* (21), 364–365 (2020)
6. Wang, L., Ding, P.: On coordination mechanism of cluster network and intelligent manufacturing technology diffusion. *J. Hangzhou Dianzi Univ. (Soc. Sci.)* **16**(1), 7–11 (2020)
7. Han, Y.: Research on the construction and practice of intelligent manufacturing specialty cluster in Chongqing higher vocational colleges. *J. Harbin Inst. Vocat. Technol.* **1**, 13–15 (2020)
8. Xiao, Y., Wang, M., Guo, L., et al.: Intelligent manufacturing plan management based on digital twins. *J. Syst. Simul.* **31**(11), 2323–2334 (2019)
9. Yu, C.: The theoretical framework for developing world-class intelligent manufacturing industry clusters and practice in Chongqing. *J. Chongqing Inst. Technol.* **33**(5), 40–48 (2019)
10. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. *Mob. Netw. Appl.* **26**(5), 1891–1903 (2021). <https://doi.org/10.1007/s11036-020-01725-x>



# 3D-HEVC Deep Video Information Hiding and Secure Transmission Method

Cai-xu Xu<sup>1(✉)</sup>, Hui Guo<sup>1</sup>, Cai-cun Cen<sup>1</sup>, and Yong-ming Chen<sup>2</sup>

<sup>1</sup> School of Electronics and Information Engineering, WuZhou University,  
WuZhou 543002, China

xuccx32@outlook.com

<sup>2</sup> Guangzhou Huali College, Guangzhou 511325, China

**Abstract.** HEVC information hiding algorithm and DCT/DST video information hiding algorithm used in the past have poor hiding effect and low security transmission efficiency. In view of this phenomenon, a new 3D-HEVC deep video information hiding and secure transmission method is designed. According to the coding principle of single depth frame, the multi-view color and depth video feature regions are divided, and then the information hiding process is designed. On this basis, the encryption key is used to realize the secure transmission of information. The experimental results show that the method has good hiding effect and high safety transmission efficiency, which proves that it has good hiding and transmission effect.

**Keywords:** 3D-HEVC deep video · Information hiding · Secure transmission · Encryption key

## 1 Introduction

3D-HEVC video can provide three-dimensional, telepresence and interactivity, so it has been widely concerned. MVD is the main representation format of 3D-HEVC video scene, which enables the decoder of 3D-HEVC video system to draw virtual viewpoint by depth map based rendering technology.

In order to compress MVD efficiently, we usually need to expand 3D on the basis of efficient video coding, and develop 3D-HEVC coding standard [1]. With the growing maturity of multimedia and network technology, 3D video acquisition and processing become easier, and the resulting information security problems become increasingly serious. Information hiding technology is an effective way to solve this problem [2].

At present, the information hiding algorithm of plane color video has been relatively mature, mainly divided into the original domain information hiding algorithm and the information hiding algorithm based on the encoding and decoding platform. In the video information hiding algorithm based on codec, secret information is directly

embedded in the video coding process. Compared with the original video information hiding method, the secret information extracted in the decoding process can greatly reduce the degradation of the decoded video quality. HEVC is the latest video coding standard. Under the same coding quality, it can save 50% bit rate compared with H.264/AVC [3]. To this end, some scholars proposed an information hiding algorithm based on HEVC efficient motion vector space coding, which can embed a certain amount of information, but has a certain influence on the overall bit rate. Another scholar proposed an HEVC video information hiding algorithm based on DCT/DST coefficient without error drift, but under the condition of high bit rate, the information hiding capacity is small.

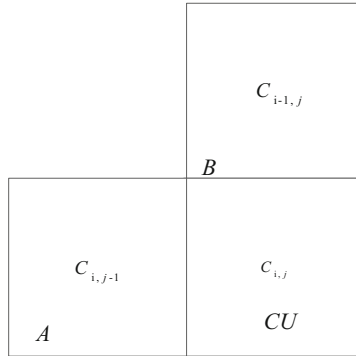
In order to safely transmit secret information in 3D video, a 3D-HEVC depth video information hiding algorithm based on multi-view video features is proposed for the latest coding and compression standards. First, map the texture feature area of the color video to the corresponding depth video, and then divide the depth video according to the edge characteristics of the depth video; then, according to the influence of different areas on the coding efficiency, the maximum coding unit (The QP value of LCU) is embedded in the secret information; finally, the modified QP value is used for encoding and compression, and the video information is transmitted.

## 2 3D-HEVC Depth Video Information Hiding Algorithm

3D-HEVC coding framework is applied to multi view video depth (MVD) system to realize video coding, decoding and multi view rendering, and provide users with real stereo vision [4]. However, in the process of multi view rendering, the accuracy of depth map boundary will directly affect the position of 3D warping, resulting in the decline of virtual view quality and user experience. From this point of view, the influence of different areas of depth video on the visual effects of virtual viewpoints is considered. Under the condition that the rendering quality and the bit rate are not changed much, the secret information is embedded in the depth map by using the characteristic of the variation of coding quantization parameters [5].

### 2.1 Coding Principle of Single Depth Intra Mode

The depth video contains a large number of flat areas, and the pixel values in the area are similar or equal. For the characteristics of depth video, if the color video coding prediction mode is still adopted, it is difficult to achieve the purpose of efficient compression [1]. Therefore, the 3D-HEVC coding standard has added a single-depth intra mode specifically for the flat area of depth video. Single-depth intra mode is a CU-level coding mode. The current CU block constructs a pixel candidate list based on the neighborhood reconstruction information, traverses the pixels in the list to calculate the



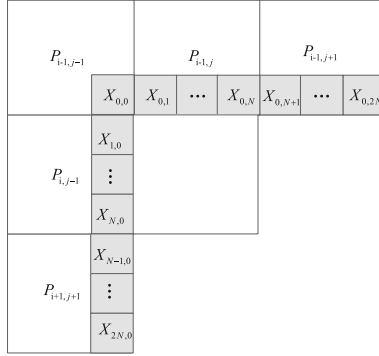
**Fig. 1.** Neighborhood information used in building candidate list

rate distortion cost, and calculates the index of the pixel with the least cost in the candidate list the value is written into the code stream without the need for transformation and quantization. The decoder directly finds the corresponding reconstructed pixel in the candidate list according to the index value to reconstruct the current block [6].

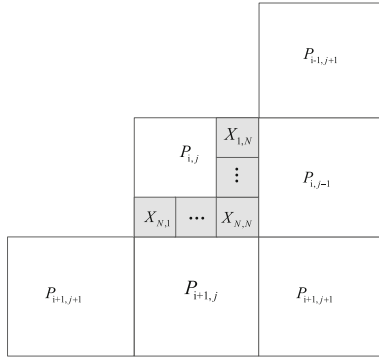
Figure 1 shows the neighborhood information used to build the candidate list. In Fig. 1, the middle pixels of the reconstructed Cu block on the left and top are A and B respectively, and the traversal order is from A to B.

In the single depth intra mode, the corresponding reconstruction pixel is selected according to the index value, and the current Cu block is reconstructed from this pixel. Therefore, when the pixel values in the candidate list are equal, using the embedding information to modulate the index value of the pixel in the current list will not cause the embedding error of the current Cu block, because the final selected reconstruction pixel has not changed [7].

In order to further improve the embedding amount of information, the situation that the pixel values of the candidate list are not equal is analyzed. The prediction reference relationship between the current Pu block and the neighborhood block is shown in Fig. 2.



(a) Reference pixel for intra prediction of current PU block



(b) The current PU block is used as a reference pixel for intra prediction in subsequent blocks

**Fig. 2.** The prediction reference relationship between the current PU block and the neighboring block

Figure 2 shows the prediction reference relationship between the current prediction unit (PU) block and the neighborhood block. Figure 2(a) is the reference pixel of the current Pu block during intra prediction. The distortion of the neighborhood pixel  $\{X_{0,i}\}_{i=0.1,\dots,N} / \{X_{i,0}\}_{i=0.1,\dots,N}$  ( $N \times N$  is the size of the Pu block) will affect the prediction of the current Pu block, resulting in error drift. Figure 2(b) is the reference pixel position of the current PU block used for prediction in the frame of the subsequent blocks. The last row and the last column of the current PU block can both be used as the reference pixel of the subsequent blocks [8].

### 2.2 Multi-viewpoint Color and Depth Video Feature Area Division

Depth video is mainly used to draw virtual viewpoint, and the edge region in depth image has a great influence on the quality of still image. At the same time, the depth video is distorted in the texture region where the corresponding position is color video, which has a greater impact on the quality of rendering virtual viewpoint [9, 10].

Considering the above factors, the depth video is divided into four categories according to the characteristics of color and depth video, namely, the color video texture region (CTR) corresponding to the depth video edge, which is called tder; the color video texture region corresponding to the depth video smooth region (DSR), which is called TDSR; and the color video flat region corresponding to the depth video edge (DER) is fder; the flat area of color video corresponding to depth video smoothing (DSR) is fdsr [11].

The specific division idea is: for depth video region, in order to improve the accuracy of depth edge division, an adaptive region edge extraction method is adopted. Firstly, the gradient value of each pixel is calculated by Sobel operator, and then the discrimination threshold is obtained adaptively through the histogram characteristics of the gradient value, and finally the Der and DSR regions are obtained. For color video region, Canny operator is used to obtain pixel gradient, and the threshold and gradient are compared to determine CTR and CFR region [12]. The specific principle is as follows:

(1) Deep video edge detection

The depth gradient value  $\nabla g(x, y)$  is obtained by using template convolution to calculate the horizontal and vertical superposition. The specific calculation formula is as follows:

$$\begin{aligned} \nabla g(x, y) &= \sqrt{g_x^2(x, y) + g_y^2(x, y)} \\ g_x &= a_x \times D \\ g_y &= a_y \times D \end{aligned} \quad (1)$$

In formula (1),  $g_x$  and  $g_y$  represent the horizontal and vertical gradient vectors of the pixel,  $D$  represents the depth value matrix, and  $a_x$  and  $a_y$  represent the horizontal and vertical gradient operators, respectively.

At present, the depth video is mainly obtained by depth estimation software, and the estimation is not accurate [13]. In order to remove the false boundary in the gradient image, the gradient histogram is used to adaptively obtain the discrimination threshold, and finally the depth boundary is detected. Assuming that there is a dual pattern distribution in the histogram distribution of depth gradient, there must be an optimal threshold to satisfy the segmentation of the depth region. By using the intra group equilibrium measure variance, an appropriate dividing line is determined to minimize the weighted sum of intra group variance.

(2) Color video texture detection

The color area division method generally uses the Canny operator. First, Gaussian smoothing filtering is performed on the color image to eliminate Gaussian noise and improve the stability of edge detection. Second, calculate the gradient value of each pixel. Then, non-maximum value suppression is performed to extract the



gradient point with the maximum extreme value as the boundary. Because the traditional single-threshold method will lose some real edges, double-threshold detection is used to select the final boundary position from the gray boundary. The calculation formula for dual threshold detection is as follows:

$$C_e(x, y) = \begin{cases} C_2(x, y) & \text{if } C_1(x, y) = 0 \cup C_2(x, y) = 255 \\ & \exists(x, y) \in N_{c_1} \\ C_1(x, y) & \text{if } C_1(x, y) = 255 \\ 0 & \text{other wise} \end{cases} \quad (2)$$

Among them,  $C_1(x, y)$  and  $C_2(x, y)$  are the edges with high threshold and low threshold to obtain non maximum suppression results, respectively. The position with a value of 255 indicates that there is an edge, and the place with a value of 0 indicates a flat area.  $N_{c_1}$  is the four neighborhood region of high threshold edge detection. The expression of color video partition is as follows:

$$\begin{cases} \text{if } C_e(x, y) = 255 & C(x, y) \subseteq \text{CTR} \\ \text{if } C_e(x, y) = 0 & C(x, y) \subseteq \text{CFR} \end{cases} \quad (3)$$

where  $C(x, y)$  represents the pixel value of the color video region corresponding to the  $(x, y)$  coordinate region.

(3) Depth video region division results

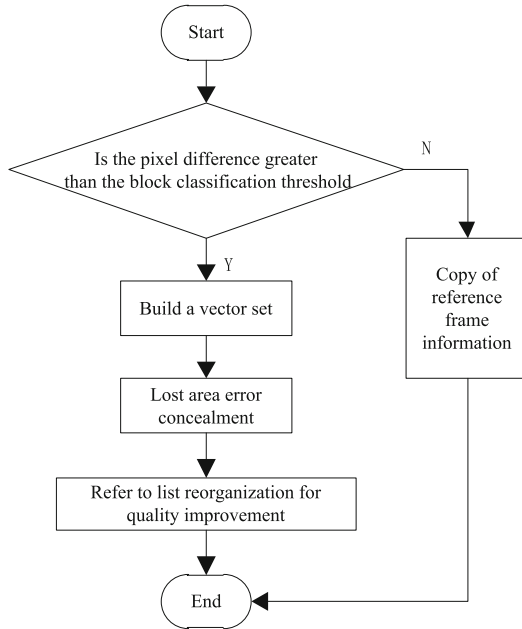
The maximum coding unit (LCU) of 3D-HEVC is the pixel block of  $64 \times 64$ . according to the characteristics of coding structure, the pixel value area of  $64 \times 64$  is divided as the basic unit. Combined with the color video region segmentation, the final depth video region segmentation is as follows

$$Rs = \begin{cases} \theta_1 & \text{CTR} \cap \text{DER} \\ \theta_2 & \text{CFR} \cap \text{DER} \\ \theta_3 & \text{CTR} \cap \text{DSR} \\ \theta_4 & \text{CFR} \cap \text{DSR} \end{cases} \quad (4)$$

Among them:  $Rs$  is the discriminant factor, which is used to determine what kind of region to divide, and the image region  $Rs$  is divided.

**2.3 Hiding Method**

The information hiding process is shown in Fig. 3.



**Fig. 3.** Flow chart of information hiding

The specific steps of the hiding method are as follows:

Step (1): according to the coordinates of the pixel points of the image missing block, the co bit block  $B_1$  in the forward reference frame and the co bit block  $B_2$  in the backward reference frame are obtained, and the size of the co bit block is  $64 \times 64$ .

Step (2): Take the absolute value of the difference between the average values of the pixels of the two co-located blocks  $B_1$  and  $B_2$ . When the average pixel difference  $Th$  is greater than the threshold 1, it is judged that the current lost block is a moving block, otherwise, it is judged as a stationary block. If it is judged that the type of the lost block is a static block, use the co-located block in the forward reference frame or the co-located block information in the backward reference frame to fill the current missing block in the image for recovery; if the type of the lost block is judged to be a moving block, Adopt the method based on vector compensation to recover; and then improve the quality of the lost blocks with poor final recovery performance.

The specific steps of restoration based on vector compensation are as follows:

Step (a): build the motion sharing vector.

After scaling the available motion vectors of the same bit block in the texture image, the set of vectors to be selected is added; if the corresponding same bit block is an intra coding block, the average value of the motion vectors of the adjacent blocks that are not zero is taken as the current motion vector to be selected:

$$MV_{\text{texture}} = \frac{MV_t}{R} \quad (5)$$

In formula (5),  $MV_{\text{texture}}$  is the final optimized motion vector,  $MV_t$  is the motion vector directly obtained in the texture map, and  $R$  is the zoom ratio.

Step (b): construct the airspace vector set.

Select the motion/disparity vectors of the lower left block 0, the left block 1, the upper left block 2, the upper block 3, the upper right block 4, and the lower block 5 of size  $4 \times 4$  to construct the spatial vector set, and the corresponding displacement vector set is  $\{SV_0, SV_1, SV_2, SV_3, SV_4, SV_5\}$ ; If the selected block is lost, the motion vector information of the block will be discarded.

Step (c): Global parallax vector.

Calculation method of global parallax vector value:

$$DV_g = \frac{1}{N} \sum_{i=0}^N DV_i \quad (6)$$

Among them,  $N$  is the number of blocks with a size of  $8 \times 8$  and disparity vectors, and  $DV_i$  is the disparity vector value of the  $i$ -th block  $8 \times 8$ ; the aforementioned motion sharing vector, 5 spatial vectors, and a global vector constitute a candidate vector set.

Step (d): Rebuild the lost block.

Select a certain candidate motion/disparity vector from the candidate vector set, and use the motion/disparity vector compensation method to fill the hole in the pixel position in the missing block;

Step (e): Calculate the matching degree of the outer boundary.

The weighted outer boundary matching algorithm calculates the matching degree  $Dobma$  between the outer boundary of the motion/disparity compensation block and the outer boundary of the lost block in the reference frame. According to the weighted outer boundary matching algorithm, the depth map outer boundary matching degree column and the texture map outer boundary matching degree are calculated, and the weighted outer boundary matching degree between the motion/parallax compensation block found in the reference frame is calculated. The weighted outer boundary image matching degree between the compensation block found with this co-located block through the texture map displacement vector; the reconstruction block with the smallest displacement vector compensation value is selected as the final filling block.

Step (f): Quality improvement reconstruction algorithm formula.

Use motion/disparity compensation to obtain candidate blocks in the reference frame as reconstruction blocks. If the missing block is a bidirectional reference, perform motion/disparity compensation in the front and rear reference frames, obtain candidate blocks in the two reference frames, and then combine the two. The pixel values of the candidate blocks are averaged as the final reconstruction block.

Step (g): Using reference frame recombination to improve the quality of lost blocks.

According to the displacement vector of the correct receiving block around the lost block, the vector compensation is used in the current position of the lost block to obtain

the reconstructed reference block in each reference frame, judge whether the reconstructed reference block position is occluded, and eliminate the unusable reference frame.

Step 1: Select occlusion judgment block.

Select  $8 \times 4 \times 4$  prediction unit PU blocks around each missing block as the occlusion judgment block  $A_1, 1 = [1, 8]$ , which are the prediction unit  $A_1, A_2$  on the missing block, the prediction unit  $A_3, A_4$  under the missing block, the left prediction unit  $A_5, A_6$  for the missing block, and the right prediction unit  $A_7, A_8$  for the missing block; Then the reference block is found by the vector compensation method, and 8 PU blocks at the same position around the reference block are selected as the occlusion judgment reference block.

Step 2: Judge whether the reference compensation block is occluded.

The reference block is found by vector compensation method, and whether the pixel value of the block increases significantly is judged by occlusion, and whether the normalized value of the pixel is greater than the judgment threshold is calculated. When one of the peripheral judgment blocks is larger than the threshold value, it is judged that the reference compensation block is occluded; if the reference compensation block is judged to be occluded, it indicates that the source frame of the reference compensation block is not suitable as the reference frame, and the reference frame caused by this part of occlusion is removed. Then, the reconstruction reference blocks obtained from each available reference frame are combined in pairs to get a new reconstruction reference block, which is used to reconstruct the damaged block; the best reconstruction block is selected as the final reconstruction block after quality improvement.

This method combines the characteristics of the new standard and improves the existing error concealment methods. It can improve the corresponding stereo video error concealment recovery technology under the new HEVC stereo video extension standard 3D-HEVC, and has a good recovery effect on the network packet loss phenomenon of stereo video in network transmission.

### 3 Information Security Transmission

In the information encryption part, mixed encryption is used. Because the DES algorithm is a 64-bit encryption for packet data, and a 56-bit key is used to encrypt 64-bit data blocks, there is no other effective method except for the exhaustive search method to attack DES., High security. In addition, the encryption and decryption processing of the DES algorithm can effectively apply the combination of binary forms. Therefore, the use of software and hardware can achieve high-speed encryption and decryption processing, and is suitable for large amounts of data processing. In terms of the difficulty of key generation, the DES algorithm only needs to generate 56 bits. In the key calculation, selection permutation and cyclic shift transformation are used in the key calculation; while the RSA method must select two secure large prime numbers  $p$  and  $q$  that are more than 100 digits different in decimal, and then calculate the public key and Private key. The speed is very slow and often does not meet the actual needs.

However, in terms of key management, the DES algorithm is inferior to the RSA algorithm. Because RSA is a public cryptosystem and uses a public form to distribute encryption keys, it is easy to update the encryption keys. And for different communication objects, you only need to keep your own decryption keys. The DES algorithm belongs to the symmetric cryptography system, which requires secret distribution of keys before communication. It is difficult to replace the keys, and for different communication objects, DES Different keys need to be generated and kept.

In the data transmission part, the traditional direct channel transmission is not used, but the information hiding method is used to hide the encrypted data in the image, video, text or audio signal before transmission. During the entire transfer process, only some unobtrusive ordinary files are exposed on the Internet. Even if intercepted by an attacker, it will be considered as meaningless files and discarded. Even if the secret information pool is extracted, the original plaintext can be obtained after repeated decryption, so the probability of data leakage is almost zero. The specific design plan is as follows:

Sender:

- (1) The sender will randomly generate the des key  $K$  for this communication, use the key  $K$  to encrypt and compress the plaintext  $m$  to get the ciphertext  $cm$ , and use the RSA public key of the system to encrypt the key  $K$  to get the ciphertext  $CK$ , and then combine the  $cm$  and  $CK$  into the ciphertext  $C$ .
- (2) When the sender encrypts the data, the ciphertext  $g$  generates a message digest by one-way hash function operation, and then encrypts the message digest with its own RSA private key to form the sender's digital signature. The sender scrambles the digital signature to get the scrambled signature  $CD$ , and combines  $CD$  and  $C$  into the ciphertext  $CC$ .
- (3) The sender uses the embedding algorithm and information sharing technology to randomly hide the scrambled signature, encrypted plaintext, and encrypted symmetric key in the carrier information to obtain the hidden carrier; then, the hidden carrier is sent to the receiver.

Receiver:

- (1) After the receiver receives the sender's information, the secret information in the hidden carrier is extracted through the extraction algorithm.
- (2) Decompose  $Cc$  into  $Cd$  and  $C$ , and the receiver restores the scrambled signature  $Cd$  to obtain the digital signature of the sender. Use the sender's RSA public key to decrypt the digital signature to get the message digest. If it can be unlocked, it means that the plaintext is indeed sent from the sender, thus verifying the authenticity of the sender's identity.
- (3) Then use a one-way hash function for  $C$  to generate a message digest. Compare the message digest and decrypt the digital signature to get the message digest. If they are the same, it means that the plaintext has not been tampered with or forged during transmission, otherwise it is false.

- (4) The pseudo receiver decomposes  $C$  into  $C_m$  and  $C_k$ ; and decrypts  $C_k$  with the RSA private key provided by the system to get the key  $K$ , and then decrypts  $C_m$  with the key  $K$  to get the plaintext  $M$ .

## 4 Experiment and Analysis

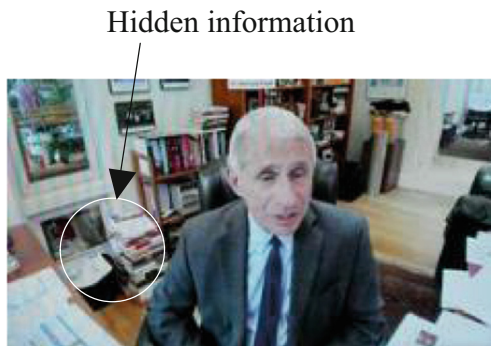
In order to verify the effectiveness of the 3D-HEVC deep video information hiding and secure transmission method, the following experiments are designed.

The experimental environment is as follows: Using the 3D-HEVC standard encoding platform HTM13.0, using the All-intra profile, encoding 100 frames, the color video encoding QP values are 25, 30, 35, and 40, and the corresponding depth video encoding QP values are 34, 39, and 40, respectively. 42 and 45, other configuration parameters are platform default values. The test sequence is 3 viewpoints and 5 viewpoints of Balloons sequence, 2 viewpoints and 4 viewpoints of Newspaper sequence, Kend. Sequence of 3 viewpoints and 5 viewpoints, Shark sequence of 1 viewpoint and 9 viewpoints, Poznan Street sequence of 3 viewpoints and 5 viewpoints, UndoDancer sequence of 1 viewpoint and 9 viewpoints, the resolution of the first 3 sequences is  $1024 \times 768$ , and the resolution of the last 3 The resolution of the sequence is  $1920 \times 1088$ .

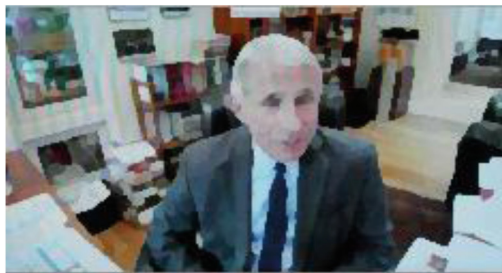
### 4.1 Video Information Hiding Verification

The imperceptibility of video is also a performance index of video information hiding algorithm, that is, there is no significant decline in video quality before and after embedding information. Because the depth video is not used for viewing directly, but for drawing virtual viewpoint, the quality change of depth video after embedding information can be reflected by the quality of drawing viewpoint. Ballons sequence, newspapersequence and shark sequence are selected to illustrate.

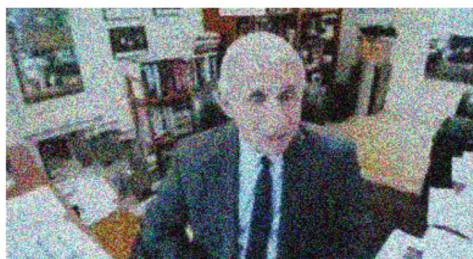
Figure 4 shows the original video sequence (Fig. 5).



**Fig. 4.** Original image



(a) HEVC information hiding algorithm



(b) DCT / DST video information hiding algorithm



(c) Multi feature points

**Fig. 5.** Comparative analysis of different hiding methods

As can be seen from Fig. 4, after the application of the method in this paper, there is no obvious distortion in the quality of the rendering view points before and after information embedding, indicating that the imperceptibility of the video is good after the application of the method in this paper.

#### **4.2 Video Information Security Transmission Verification**

In order to further verify the application performance of the method presented in this paper, the following comparative experiments were designed: the method presented in this paper, the traditional HEVC information hiding method and the DCT/DST video

information hiding method were respectively used to verify the secure transmission efficiency of video information. The results are shown in Table 1.

**Table 1.** Three methods of video information security transmission efficiency

Transmission times/time	HEVC information hiding method	DCT/DST video information hiding method	This paper method
1	0.62	0.72	0.98
2	0.58	0.75	0.99
3	0.55	0.83	0.97
4	0.60	0.85	0.98
5	0.57	0.86	0.98

The results of Table 1 show that the highest transmission efficiency of video information is 0.99 after using this method, which shows that this method has good transmission effect. After the application of the two traditional methods, the transmission efficiency of video information is always below 0.9. By comparison, we can see the advantages of this method.

## 5 Conclusion

In this study, a 3D-HEVC deep video information hiding and secure transmission method is designed. The method combining with the video image texture feature and edge information of depth video area, then according to the different areas of the depth video on the quality of the rendering virtual view, adopt different ways of modulation maximum coding unit (LCU) QP embedded secret information, on the basis of designing the concrete methods of information hiding and secure transport. According to the experimental results, this method not only has a large embedding capacity, but also can better guarantee the subjective and objective quality of video, and meets the requirements of invisibility, security and real-time of information hiding.

In the following research, we will consider to complete the information hiding processing in the process of video production, so as to effectively ensure the quality of the rendering view before and after the deep video embedding information without distortion.

**Fund Projects.** National Natural Science Foundation of China: The Key Technologies about Fast Coding and Quality Controlling of Fractal Image Compression (61961036).

Natural Science Foundation of Guangxi: Research on the Key Technologies about Decoder for Reliable Transmission of HEVC for Microscopic Video(2020JJA170007).

Industry and Education and Research projects for Wuzhou University and Wuzhou High-tech zone: the portable remote control and HD Endoscope Camera System and cloud service platform Construction(2020G001).

Basic Ability Improvement Project for Young and Middle-aged Teachers in Guangxi: Research on fast image retrieval algorithm based on fractal coding (2018KY0537).



Wuzhou Science and Technology Development Project: The Research and development of the visual auto-generating platform of VR and demonstration of application of construction machinery (2020B02003).

Natural Science Foundation of Wuzhou University: Research on the Key Technologies of integrated navigation and positioning for GNSS receiver (2020C001).

## References

1. Ma, L.V., Yu, G., Kim, J.Y., et al.: An efficient transmission method based on HEVC multi-view adaptive video streaming over P2P network in NFV. *J. Supercomput.* **74**(12), 6939–6959 (2018)
2. Hossain, M.S., Muhammad, G., Abdul, W., et al.: Cloud-assisted secure video transmission and sharing framework for smart cities. *Futur. Gener. Comput. Syst.* **83**(1), 596–606 (2017)
3. Kitagawa, W., Inba, A., et al.: Objective function optimization for electrical machine by using multi-objective genetic programming and display method of its results. *IEEJ Trans. Electron. Inf. Syst.* **139**(7), 796–801 (2019)
4. Wang, F., Qi, H., Zhou, X., et al.: Demonstration programming and optimization method of cooperative robot based on multi-source information fusion. *Jiqiren/Robot* **40**(4), 551–559 (2018)
5. Xu, G., Chen, F., Li, X.T., et al.: Closed-loop solution method of active vision reconstruction via a 3D reference and an external camera. *Appl. Opt.* **58**(29), 8092–8100 (2019)
6. Cao, X., Xuan, S., Hu, T., et al.: 3D printing-assistant method for magneto-active pulse pump: experiment, simulation, and deformation theory. *Appl. Phys. Lett.* **117**(24), 241901 (2020)
7. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Compl. Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
8. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. *Entropy* **21**(9), 902–912 (2019)
9. Liu, S., Bai, W., Zeng, N., et al.: A fast fractal based compression for MRI images. *IEEE Access* **7**, 62412–62420 (2019)
10. Zheng, Y.: Formal process virtual machine for smart contracts verification. *Int. J. Performab. Eng.* **14**(8), 1–9 (2018)
11. Zhang, T., Liu, Y.N., Xing, Y.L., et al.: Lossless information hiding in AMBTC domain based on histogram shift. *Appl. Res. Comput.* **36**(6), 1771–1775 (2019)
12. Ren, S., Wang, Z., Su, D.X., et al.: Information hiding algorithm based on mapping and structure data of 3D model. *J. Commun.* **40**(5), 211–222 (2019)
13. Cui, B.D., Xin, C., Pei, X.S., et al.: A security reversible information hiding algorithm based on histogram shifting. *Sci. Technol. Eng.* **19**(22), 215–222 (2019)



# Research on Difference Elimination Method Between Small Sample Databases Based on Feature Extraction

Jin-hua Liu<sup>1</sup> and Fu-lian Zhong<sup>2(✉)</sup>

<sup>1</sup> Department of Professional and Continuing Education, XinYu University, XinYu 338031, China

<sup>2</sup> School of Mathematics and Computer Science, XinYu University, XinYu 338031, China  
zhongfulian682@163.com

**Abstract.** The traditional method of eliminating the differences between small sample databases takes a long time and has a low accuracy. Therefore, a method of eliminating the differences between small sample databases based on feature extraction is designed. In order to realize the data communication between small sample databases, we construct the data retention mechanism of small sample databases, store the sample data safely, discretize the data attributes, sort the primary and secondary relationship of the sample data, select the optimal integration and sharing path of the sample data, cluster the sample data, and select the cluster head and relay node. Eliminate the differences between small sample databases. The experimental results show that compared with the traditional method, this design method shortens the time of eliminating the differences between small sample databases, and improves the accuracy of eliminating the differences.

**Keyword:** Small sample database · Data clustering · Data fusion · Difference elimination

## 1 Introduction

With the rapid development of Internet, a large number of users access the data and information resources in small sample database through Internet. The characteristics of small sample database are that there are differences in logic and physics. Using different information resources to operate each other in small sample database has become a hot research topic in database field. The difference elimination of small sample database is to access the database transparently in the computer network environment, support the operation between databases, and update, merge and sort the other database based on the support of database application tools. In order to eliminate the differences between small sample databases, it is necessary to shield the differences between databases distributed in different sites. In the network environment, the differences between databases have two aspects: on the one hand, the database hardware and software environment is different; On the other hand, there are differences in the database itself, including the differences caused by data semantics.

In the anti underwater cooperative communication system, nodes can send data, and nodes can share the uplink and downlink frequencies, so as to realize the reuse of database resources, Eliminate differences between databases. The domestic research on the methods of eliminating the differences between databases has also made great progress. This paper proposes a method of eliminating the differences between small sample databases based on KL divergence. KL divergence is used to analyze the degree of differences between databases, and the training data is used to generate a classifier to classify the data in the database, and the results are combined to form a data space, which is measured by complementary information entropy. This paper classifies the uncertain information in the database, judges the differences between databases, and uses the basic criteria of spatial differences to eliminate the differences between databases by simulating data sets. However, the traditional method of eliminating the differences between small sample databases takes a long time and can not eliminate the differences between databases correctly.

In order to solve the shortcomings of traditional methods, this paper designs a method to eliminate the differences between small sample databases based on feature extraction to ensure the efficiency and accuracy of eliminating the differences between databases. The main work of this paper is as follows:

- (1) The data retention mechanism of small sample database is constructed to store the sample data safely, and the data attributes are discretized to order the primary and secondary relationship of the sample data.
- (2) Select the optimal integration and sharing path of the sample data. On the basis of clustering the sample data, select the cluster head and the relay node to realize the data communication among small sample databases.
- (3) Extract model training data by features, and eliminate differences among small sample databases by fusion of communication sample data.

## **2 Design of Difference Elimination Method Between Small Sample Databases Based on Feature Extraction**

### **2.1 Building Data Retention Mechanism of Small Sample Database**

Based on the principle of cryptography, the sample data is encoded, and the data retention mechanism of small sample database is constructed. In the sample data storage process, the random dynamic time-varying factors are added to control the data storage state in real time. The data retention mechanism is divided into two parts: upload and download, and pseudo transformation. In the upload and download process, the implementation of cryptography coding based on cryptography principle is used to dynamically repair the network data, and tolerate some invalid network nodes [1]. For the data upload stage, the sample data is segmented to get multiple data blocks of the same size. The random linear coding method is used to store the coded data blocks to nodes. For the data read stage, the coding blocks are downloaded randomly to merge the sample data. For the stage of mimicry transformation, the less functional storage regeneration code is used to screen out the network nodes that need to be repaired,

control the repair bandwidth, code randomly and construct new data blocks to obtain the data nodes after mimicry transformation, and replace the original data content to ensure the integrity of sample data. The calculation formula of repair bandwidth  $Q$  of storage regeneration code is as follows:

$$Q = \frac{\xi\beta(K-1)}{L(K-L)} \tag{1}$$

Among them,  $L$  is the number of fixed size original blocks,  $\beta$  is the number of coding blocks to store the regeneration code,  $K$  is the number of network data nodes, and  $\xi$  is the repair bandwidth cost [2]. The amount of data stored in the network node is taken as the minimum storage extremum point of the regeneration code. After the node is repaired, the pseudo transformation parameters are negotiated to update the storage content of the sample data, so as to ensure that the stored data is always in the coding state, so as to download all the data on the network node. The data coding network structure is shown in Fig. 1.

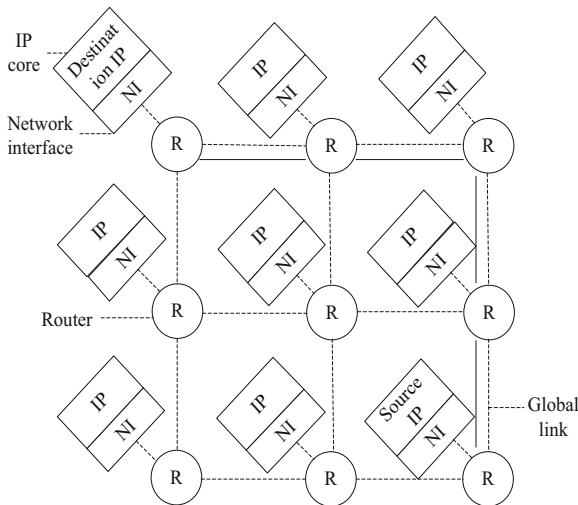


Fig. 1. Data coding network structure

As shown in Fig. 1, the data coding communication structure is composed of link, IP core, router and IP core. The network interface is used to connect the router and IP core, transmit the source IP packets to the destination IP, and use the router to determine the transmission direction. The router is composed of control logic, buffer and crossbar. The control logic is used to arbitrate the channel, calculate the route, and determine the transmission direction. The routing decision supports the authorization of the sample data, which is then transmitted to the next router through the crossbar until it reaches the packet destination.

In order to reduce the security threat of the database network environment, a data retention mechanism is constructed after data access and storage by encoding. On the basis of analyzing the repeated organization form of encoded data, the encoded data files are copied for many times, and the file copies are allocated so that they can be transferred to different storage nodes. At the same time, the file is divided into several subfiles, the file is encoded into redundant data blocks, and the low-density check code is configured for each data block. The original file of the database is reconstructed through any data block in the redundant data, so as to avoid the leakage of sample data and part of the original file when there is network attack, so as to ensure the security of the sample data [3]. The implementation process of the retention mechanism is shown in Fig. 2.

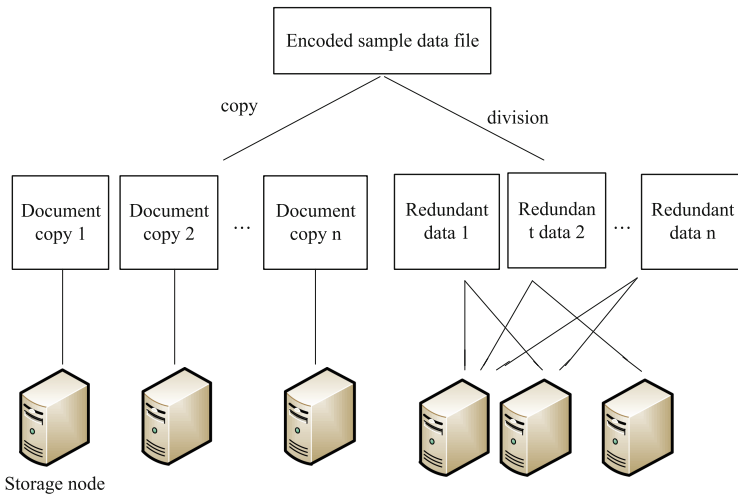


Fig. 2. Implementation process of data retention mechanism

As shown in Fig. 2, combined with storage technology and network communication technology, the storage nodes distributed in small sample databases are connected to establish a logical unified storage server. The expression of redundancy process of coded data is as follows

$$G_{n \times l} = \begin{bmatrix} I_{k \times k} \\ \phi_{k \times (n-k)} \end{bmatrix} O_{k \times l} \tag{2}$$

Among them,  $G_{n \times l}$  is the  $n$  redundant data blocks divided by  $k$  block file,  $O_{k \times l}$  is the original data column vector composed of  $k$  original data blocks,  $l$  is the number of original data block columns,  $I_{k \times k}$  is the  $k \times k$  order identity matrix, and  $\phi_{k \times (n-k)}$  is the

$k \times (n - k)$  order Cauchy matrix. In the segmented redundant data, the formation vector  $G'_{k \times k}$  of  $k$  redundant data blocks is arbitrarily extracted:

$$G'_{k \times l} = U_{k \times k} G_{n \times l} \quad (3)$$

where  $U_{k \times k}$  is a Vandermonde matrix of order  $k \times k$ . Then the formula for obtaining the original data of the data retention mechanism is as follows:

$$O_{k \times l} = U_{k \times k}^{-1} G'_{k \times l} \quad (4)$$

Through the formula (4) to complete the decoding and decryption operation, restore the original file of the small sample database, restore the original data content. So far, the construction of data retention mechanism of small sample database is completed.

## 2.2 Preprocessing Sample Data of Small Sample Database

Preprocess the sample data in the database, discretize the continuous numerical attributes, and sort the primary and secondary relationship of the sample data. The maximum minimum normalization formula is used to transform the original data linearly

$$V = \beta \frac{(L-M)}{(M-N)} \quad (5)$$

Among them,  $L$  is the data value,  $M$  and  $N$  are the maximum and minimum values of the original data of the same attribute,  $\beta$  is the mapping interval, and  $V$  is the mapped value of the original data. Using the clustering function of neural network, using the network center instead of the continuous value of the original data, clustering the continuous attributes of the mapping data, transforming the data attributes into discrete values, ensuring the relative relationship of the attributes, displaying regular rules, reducing the number of values of the same attribute data [4]. The radial basis function of Multivariable Interpolation is used in the neural network, and the three-layer forward neural network is selected as the typical structure of the neural network. In the middle layer, the attribute features extracted from the input layer are transformed to make the data category closer to the center of the network. The specific training process is shown in Fig. 3.

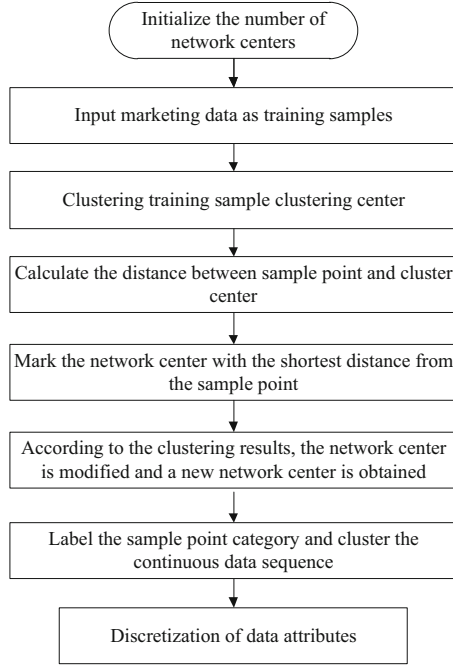


Fig. 3. Training process of database sample data

Suppose that the output value of the  $i$  neuron is  $x_i$ , the sample point of the  $j$  network center is  $G_j$ , and the middle layer neuron of the  $j$  network center is  $T_j$ , then the modified new network center  $B$  is:

$$B = \frac{\sum x_i}{G_j} \quad \forall x_i \in T_j \tag{6}$$

The mapping data sample points are divided to the new network center  $B$ , and the network center set is used as the value range instead of the sample point value, so as to eliminate the different dimensions of each dimension data, so as to find the change rule between the database sample data. Using sprint classification algorithm, sort the primary and secondary relationship of database sample data. In the sample data, select the attribute with the highest priority as the root, provide the preprocessed attribute set, search the commonness from the massive sample data, make a series of decisions, classify the sample data, split the decision tree node, and then split the sample data attribute, so that the attribute is accurately associated with the child node, and get the attribute value segmentation training set [5]. If the number of data set categories is  $m$  and the number of data set categories is the number of leaf node categories, the calculation formula of splitting parameter  $F$  is as follows:

$$F = 1 - \sum_I^m p_I^2 \quad (7)$$

where  $p_I$  is the relative frequency of data set category  $I$ . In the training set, a data node is selected to represent the data of the training set as the internal node of the decision tree. The branch structure of logical judgment is used as the edge of the decision tree, and the data attributes are associated with the root node of the decision tree to construct a multi tree decision tree. When all marketing data belong to the same category, the class label is used to define the leaf node. When the sample data does not belong to the same category, the data attribute is measured according to the information entropy, and the data in the original attribute set is deleted. When the candidate set is empty, the leaf node is returned and marked as a common category [6–8]. For different types of sample data, the calculation formula of information entropy  $Q$  is as follows:

$$Q = - \sum_{I=1}^m \frac{|C_I|}{|\xi|} \log_2 \left( \frac{|C_I|}{|\xi|} \right) \quad (8)$$

where  $\xi$  is the given training set of decision tree and  $C_I$  is the set of data set belonging to class  $I$  objects. The training set  $\xi$  is classified according to the characteristics of attributes to obtain multiple different objects. The weighted sum of the information entropy  $Q$  is carried out through the partition entropy to calculate the information gain attributes of marketing data

$$A = - \sum_{I=1}^{\phi} \left( \frac{|\zeta|}{|D|} \times Q \right) \quad (9)$$

Among them,  $A$  is the gain of marketing data information,  $\phi$  is the attribute characteristic quantity of training set,  $D$  is the information quantity required for training set classification, and  $\zeta$  is the information quantity required by training set partition. In the attribute set, select the attribute with the highest information gain  $A$ , mark the leaf node, get a score of the highest information gain attribute, and make the training set subset elements meet the score value. When the categories are the same at the node, the remaining attributes cannot be subdivided, or the given score value has no data, the class label is created, the decision tree partition is terminated, and the classification of database sample data is completed. At this point, the preprocessing of database sample data is completed.



### 2.3 Choose the Best Integration and Sharing Path of Sample Data

For the preprocessed sample data, select the optimal integration and sharing path, that is, the link path of the sample data, and add and modify the sample data. Firstly, the associated semantics of the sample data is standardized, and the frequency of different semantic query words is counted, from which the core query words are determined, and the four attributes of the core query words are determined, and then the similarity between different sample data and the words is calculated. If the similarity distance formula is adopted, the calculation formula of similarity  $Q$  is as follows:

$$Q = \left( \sum_{i=1}^4 (a_k - a_j)^k \right)^{\frac{1}{k}} \tag{10}$$

Among them,  $i$  is the four attributes of the core query word,  $a$  is the associated data of the database sample data, and  $k, j$  is the visual spatial dimension and the quantitative value of similarity distance of the data, where  $j$  is 2 or 3, when  $k$  is 1, it represents the real distance between the core query word and the spatial latitude, and when  $k \neq 1$ , it is the exact distance, representing the sum of absolute wheelbase on the spatial dimension [9]. By transforming formula (10), we get the optimal path  $S$  of transformation infimum distance in visual space dimension:

$$S = \frac{1}{cQ} \tag{11}$$

where,  $c$  is the frequency of database sample data and core query words. When  $0 < c < 1$ , the optimal path  $S$  value is between (0, 1). The smaller  $c$ , the closer  $S$  value is to 0. When  $f c > 1$ , the smaller  $c$ , the closer  $S$  value is to 1. According to the frequency of different sample data, the values of  $k, j$  and  $K$  are determined to make  $S$  reach the limit value, and the final path of data transformation is obtained. By using this path, the associated data of sample data is transformed, and the optimal mode of resource integration is obtained. Remove the query words that are not related to the associated semantics, improve the service function of the associated data, and according to the unified specification of the small sample database, map the associated data and resource ontology to form the integrated data of the small sample database [10, 11]. Calculate the sample data sharing path, let the other database download ciphertext, decrypt to obtain the complete sample data. Multiple wireless links are set up in the transmission range. Minimizing the average energy consumption of data transmission is taken as the target of channel selection. The time is divided into fixed slot periods, so that the database of the opposite party can generate sample data at a given rate, fix the transmission power of the provider, and detect the channel every other period. The channel detection process is shown in Fig. 4.

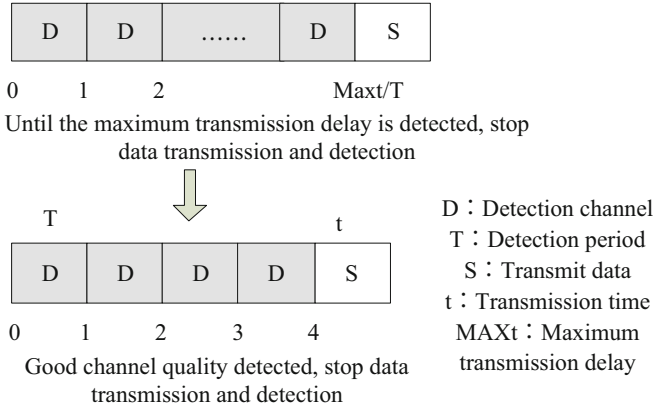


Fig. 4. One round channel detection process

The detection duration of the control channel is far less than the detection period, so that the transmission time of the database sample data can meet the requirements of the transmission delay. All the detected good channels are filtered and the channel transmission rate  $R$  is calculated:

$$R = H \log_2 \left( 1 + \frac{\phi P}{\xi H} \right) \tag{12}$$

Among them,  $\phi$  is channel gain,  $P$  is power of energy data transmission,  $H$  is channel bandwidth, and  $\xi$  is channel power spectral density. According to formula (12), transmission rate is proportional to channel gain, so real-time detection of transmission rate of each channel, capture the time with large channel gain value, obtain the channel of optimal transmission time, thus improving the transmission amount of energy data and reducing the average energy consumption of unit data. The channel is used as a sample data sharing link, and the ciphertext data is sent to the shared link, and the block chain is broadcast throughout the network, so that it is transmitted from the block head of the blockchain to the public key address of the other database. When the user receives the ciphertext, the private key  $(B, c)$  matching the public key  $(B, a)$  is selected. The formula of  $G$  of matching algorithm is:

$$G = \frac{\beta \prod_{i \in N} (e(B, a)e(B, c))^{\varpi_i}}{e(B, a)} \tag{13}$$

Among them,  $\beta$  is the mapping order of the keyword index tree,  $N$  is the set of ciphertext index structures,  $i$  is the smallest subset of index structures,  $\varpi_i$  is the subset hidden matrix,  $e(B, a)$  and  $e(B, c)$  are the attribute value sets of public key and private key respectively. After the information data is downloaded, the energy data encrypted by the public key is decrypted by using the matched private key, so that the other

database can interpret the shared sample data. So far, the optimal integration and sharing path of sample data is selected.

## 2.4 Sample Data Transmitted by Clustering Database

The cluster label of the small sample database is consistent, the cluster center of the database is output, and the transmitted sample data is clustered. Firstly, the eigenvalue of database is calculated, and each data object is regarded as a cluster class by using the classification hierarchy algorithm. Through the iterative cycle, the adjacent cluster classes are merged until all the cluster classes are merged into one cluster class. According to the specific similarity function and neighborhood threshold, the distribution of the sample data in the whole space is calculated, the known data sets are decomposed hierarchically, the high-dimensional observation sample data sets are found, the maximum eigenvalue and minimum eigenvalue of the big data are calculated, and the intrinsic geometric structure of the high-dimensional sample data is obtained. For a given small sample database, the support degree  $L_i$  of itemset  $i$  is defined:

$$L_i = \frac{|k_i|}{l} \quad (14)$$

Among them,  $k_i$  is the number of occurrence of item set  $i$  in small sample database and  $l$  is the number of local databases in small sample database.  $L_i$  is the characteristic information of small sample database. If the frequent item sets of two small sample databases have more common elements, the similarity between small sample databases is relatively large. Select the database with close support degree  $S_i$  of item set, calculate the similarity of feature value, use iterative relocation technology, redistribute the category of data objects, and determine the probability of classification of small sample database, The concept of link in clustering process of small sample database is obtained. Using ADMM algorithm, the distinguishing features are selected, the local geometry is linked to the database, the cross-correlation matrix is used to update the sampling probability of each small sample database, the cluster quality value of data samples is transmitted, the frequent sequence pattern in the small sample database is mined, the sequence prefix in the small sample database is analyzed, and the frequent term set is projected one by one. The variable data of feature values are standardized by using the same dimension minner distance, and the data close to the distance are classified into the same category. The sample data in local database is processed in a standard way. The two-dimensional logical table of database is used to record all fields and analyze the hidden structure of sample data, so that all hidden variables can correspond with data clustering, and ensure the integrity and consistency of the label of sample data cluster.

Input the small sample database for clustering analysis, use genetic algorithm to get the sample data of each database, get the sample data and the cluster to which the sample belongs, get the sum of the dimensions of the sample data of different clusters, get the center value of genetic operation, get the distance between the sample data and  $K$  clusters, and collect the small sample database with high similarity. The distance

between clusters is calculated. The data set with multiple sample data is divided into  $k$  classes of K-means clustering algorithm. The distance from the center of each cluster to the center of the whole domain is taken as the distance between clusters to calculate the distance within clusters. If the distance within the cluster is smaller, the closer the distance of the same cluster data is, the better the clustering effect is; If the distance within the cluster is larger, the distance between the data in the same cluster is larger, and the clustering cohesion is smaller. According to the classification utility of the clustering results, a new node is created as the clustering center of the small sample database, which divides the clustering groups, processes the low dimensional sample data by clustering, constructs a multi-dimensional low dimensional space, locally embeds the dimension of the sample data, reduces the dimension of the original sample data, realizes the dimension reduction of the sample data, and obtains the low dimensional representation of the small sample database.

On this basis, the similarity between data samples is taken as the objective function of the original data samples, and the sample data is embedded into the high-dimensional space to reveal the potential structure of the sample data. The small sample database is decomposed by matrix decomposition and sparse coding, and the matrix factors are non negative constrained to ensure the neighborhood of each sample data, which conforms to the non negative performance of visual data [12]. In each iteration of the small sample database, the weight of clustering partition is used to measure the quality of clustering partition, the sampling probability distribution of data samples is calculated, new sample data is randomly sampled, new K-means clustering partition is generated, and multiple clustering results are gathered, The small sample database after clustering is obtained. So far, the clustering of sample data in database transmission is completed.

## 2.5 Select Cluster Head and Relay Node for Sample Data Clustering

For the clustered database sample data, the cluster head and relay node are selected to realize the data communication between small sample databases.

The energy consumption of cluster head node includes collecting, fusing and forwarding the data of the node members in the cluster, and the energy consumption is large. The selection of cluster head is an important factor to ensure the elimination of differences between databases. Select a relay node in each cluster member, use the cluster head to collect the data from the members, analyze and process the sample data, and send the processed data to the relay node. The relay node is mainly responsible for data transmission, sharing the communication work between the cluster head and the base station. Cluster heads consume more energy when they receive their members' energy. When the cluster head communicates with the base station, the energy consumption is greatly reduced, and the relay node can be used as the relay data transmission. When communicating with the remote base station, the energy consumption of other parts is reduced, which plays a role in balancing the network energy load [13].

When the cluster head is selected, the topology of the cluster can be determined, and the relay node can be selected according to the application requirements in the

cluster. The distance between the relay node and the cluster head and the distance between the relay node and the base station can be calculated respectively. According to the minimum value of the sum of distances, the relay node is selected. After selecting the cluster head and relay node, the energy consumption of cluster members transmitting data to the cluster head and the energy consumption of the cluster head sending data to the relay node and the relay node sending data to the base station are calculated respectively to provide energy for data communication between small sample databases. So far, the selection of cluster head and relay node is completed.

## 2.6 Eliminating Differences Between Small Sample Databases Based on Feature Extraction

In the communication process of small sample database, the feature extraction model is used to fuse sample data, and then the differences between small sample databases are eliminated.

Firstly, backup the differences between small sample databases. The process is: to block data files, generate verification sum data files, search for matching blocks in small sample database according to the length of data files and data file blocks, obtain differential data files according to search results, and transfer unmatched data blocks to backup center. Assume that data blocks need to increase the number of shared blocks in the backup center. Then, you need to assign a unique block number to get the instruction file. The backup center data file difference, data file differential backup process is: according to the instruction file, get the modified data file, the backup center obtains the data file sent by the client, reconstruct a data block, and construct the final instruction file according to the check and the length of the modified data block. Complete the differential backup of the data blocks in the small sample mode.

After the backup of the differences between small sample databases, the feature extraction model is used to fuse the node data in the database, and the training of the feature extraction model is completed. The difference elimination between small sample databases is realized by using data fusion.

The feature extraction model uses machine learning convolution neural network to make the output number of features equal to the number of convolution cores. The convolution kernel of fixed number is used to convolute the sample data to ensure convolution extraction features. Only the cognitive area of convolution core is included. In the convolution process, the edge features of the bottom layer are extracted first, and then convolution operation is used. Further, the feature extraction of the underlying features is carried out. After repeated operation, the high-level features with abstract significance are obtained.

In order to control the sample data output accurately, the latter layer of neurons can be connected with the local features of the neural network front layer, ensure that all connection information is summarized in the high-level network, strengthen the relevance of the database sample data, reduce the network parameters of local connection, and obtain the global information of the sample data. During the convolution core movement, the data can be obtained, The inner parameters of convolution core are fixed, and the convolution core of convolution layer is input by weight sharing method,

so as to reduce the parameter of neural network. The convolution expression formula of feature extraction model is as follows:

$$A_{n,m} = \sum_{v=1}^V \sum_{u=1}^U w_{v,u} \times a_{i+v,j+u} + b \quad (15)$$

Among them:  $A_{n,m}$  is the characteristic element of the convolution output of the cognitive region of the sample data in row  $n$  and column  $m$ ;  $w_{v,u}$  is the convolution kernel weight parameter of the machine learning neural network in row  $v$  and column  $u$ ;  $V$  is the width of the convolution kernel of the neural network;  $U$  is the height of the convolution kernel of the neural network;  $a_{i+v,j+u}$  is the correlation semantics of the  $i+v$  and  $j+u$  columns of the sample data;  $i$  and  $j$  are the  $i$  and  $j$  columns of the convolution kernel sliding in the cognitive region.

After the convolution layer is completed, the pooling layer is used to retain the main features of the output feature elements, enhance the anti-interference of the sample data features, and retain more feature data. Through the activation layer, the feature data learned by the machine is mapped to the label space, and the feature data of the small sample database is received comprehensively. The cluster head sends the trained feature data to the terminal node, fuses the data in the database, and sends the fused data to the cluster head node, and classifies the data. Data fusion is used to eliminate the differences between databases, which can reduce the amount of data and reduce the energy consumption in the process of database difference elimination, So as to realize the data collection, fusion and transmission between small sample databases.

So far, we have completed the design of the difference elimination method between small sample databases based on feature extraction.

### 3 Experiment and Analysis

The design method is compared with the two groups of traditional methods to eliminate the differences between small sample databases, and the efficiency and accuracy of eliminating the differences between small sample databases are compared.

#### 3.1 Experimental Preparation

The test environment of the experiment is: 2G memory client, OS is Windows 2008 R2 500G disk, CPU is IntelCore™ 2 Duo17500, Microsoft SQL 2008 R2 is installed in the client, and database is created in the client, which is developed by object-oriented Java language. The created small sample database includes Oracle, SOL Server2000, SQLServer2005, MYSOL, Access, DB2, XML, COIL20 UMIST, FERET, PIE, BA, ISOLET1. Small sample databases are of the same type and are distinguished by different table spaces. The specific distribution is shown in Table 1.

**Table 1.** Deployment of small sample database

Database type	Database distribution	Database name
Oracle	16.200.11	TIAGT
SOLServer2000	16.200.11	HAL
sQLServer2005	16.200.11	LE
MySQL	16.200.11	UAIDLL
Access	16.111.25	UIPJ
DB2	16.111.25	UWTAHJ
XML	16.111.46	nlabdcl
COIL20	16.110.36	nlabdc2
UMIST	16.111.18	simu2
FERET	16.200.34	VMcine
PIE	16.110.25	CZGT
BA	16.110.25	VRDisp
ISOLET1	16.110.17	GIS

A distributed parallel computing environment is built. Five computers are set up in the LAN to form a Ha-doop cluster. One computer is defined as a Master node, and the other computer is defined as a Slave node. All computers have 2 GB memory and 3.20 GHz. Thirteen types of databases and one FTP file server are distributed in the LAN. The sample data of the database is shown in Table 2.

**Table 2.** Database summary sample data

Database type	Number of samples	Characteristic number	Number of categories
Oracle	1448	1022	22
SOLServer2000	571	645	291
sQLServer2005	1470	1290	281
MySQL	1422	1021	278
Access	1407	325	86
DB2	152	602	20
XML	1491	1217	291
COIL20	114	1008	129
UMIST	1091	437	198
FERET	902	572	116
PIE	1210	209	402
BA	1002	319	273
ISOLET1	1472	1934	73

### 3.2 Experimental Result

#### Experimental Results on the Efficiency of Eliminating Differences Between Databases

The design method and two traditional methods are used to record the time cost of difference elimination between small sample databases. Set the number of iterations to 20, change the network bandwidth of the LAN, and the comparison results of elimination efficiency are shown in Fig. 5.

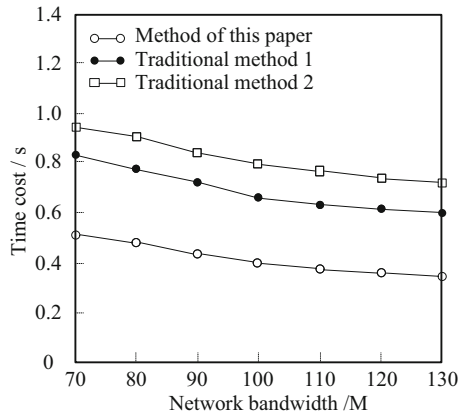


Fig. 5. Elimination time under different network bandwidth

Set the network bandwidth of the local area network to 100 m, change the number of iterations, and the comparison results of the elimination efficiency of the three groups of methods are shown in Fig. 6.

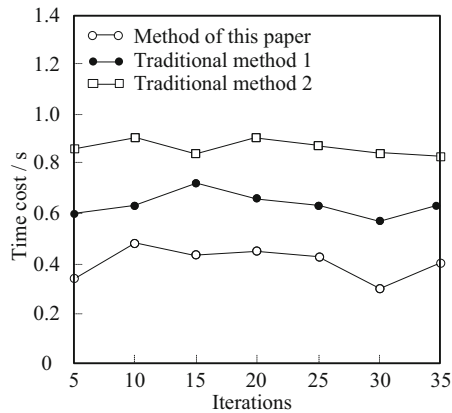


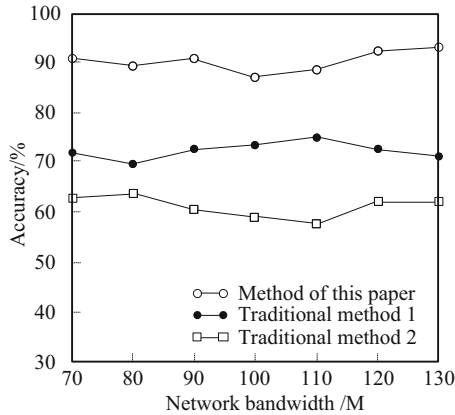
Fig. 6. Elimination time under different iterations



As can be seen from Fig. 6, the elimination time is always less than 0.5 s after the application of the method in this paper. After the application of traditional method 1, the elimination time was between 0.6 s and 0.75 s. After the application of traditional method 2, the elimination time was always above 0.8 s. In summary, under different network bandwidths and different iteration times, it takes time to eliminate the differences among the small sample databases of the method in this paper. It is obviously lower than the two sets of traditional methods, which proves that it can improve the efficiency of eliminating the differences between databases.

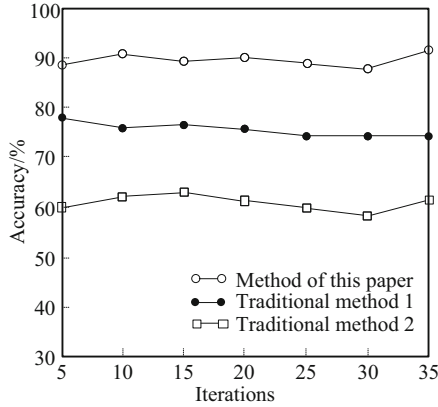
**Experimental Results of the Accuracy of Eliminating the Difference Between Databases**

The design method and two traditional methods are used to test the correctness of difference elimination between small sample databases. Set the number of iterations to 20, change the network bandwidth of the LAN, and eliminate the error rate. The comparison result is shown in Fig. 7.



**Fig. 7.** Elimination accuracy under different network bandwidth

Set the network bandwidth of the local area network to 100 m, change the number of iterations, and the comparison results of the elimination accuracy of the three methods are shown in Fig. 8.



**Fig. 8.** Accuracy of elimination under different iterations

As can be seen from Fig. 8, under different network bandwidths and iterations, the accuracy of difference elimination among small sample databases is around 90% after the application of the method in this paper. The accuracy of traditional method 1 is less than 80%, and that of traditional method 2 is less than 65%. Therefore, the accuracy of the proposed method in eliminating the differences between small sample databases is significantly higher than that of the two groups of traditional methods, which proves that the proposed method can correctly eliminate the differences between databases.

## 4 Conclusion

This paper designs a method to eliminate the differences between small sample databases based on feature extraction. After securely storing the sample data, the data attributes are discretized to sort the primary and secondary relationships of the sample data. Then select the optimal integration and sharing path of sample data. On the basis of clustering sample data, select the cluster head and the relay node to realize the data communication among small sample databases. Then feature extraction model training data is used to eliminate the differences among small sample databases by fusion of communication sample data. This method gives full play to the technical advantages of feature extraction, shortens the time consuming of eliminating differences among small sample databases, and fully guarantees the correctness of eliminating differences among small sample databases. However, there are still some shortcomings in this study. In future studies, subject analysis and data mining functions will be added to the database to provide favorable conditions for data integration of small sample databases.

## References

1. Guo, Y., Zuo, J.: DAO pattern database elimination simulation based on big data analysis. *Comput. Simul.* **36**(12), 336–340 (2019)
2. Zhang, J., Wang, R., Jiang, X., et al.: Research and implementation of synchronization technology of heterogeneous database. *Software Eng.* **24**(1), 6–9+5 (2021)
3. She, J., Guo, Y.: Design of virtual database resource reorganization system based on multimedia technology. *Modern Electron. Technique* **44**(2), 86–90 (2021)
4. Licheng, L.I.U., Yifan, X.U., Guicai, X.I.E., et al.: Outlier detection and semantic disambiguation of JSON document for NoSQL database. *Comput. Sci.* **48**(2), 93–99 (2021)
5. Shen, D., Yang, G.: Heterogeneous database integration middleware system for privacy protection. *Comput. Technol. Dev.* **30**(1), 99–105 (2020)
6. Liu, S., Li, Z., Zhang, Y., Cheng, X.: Introduction of key problems in long-distance learning and training. *Mob. Networks Appl.* **24**(1), 1–4 (2018). <https://doi.org/10.1007/s11036-018-1136-6>
7. Liu, S., Sun, G., Fu, W.: *e-Learning, e-Education, and Online Training*, pp. 1–386. Springer, Cham (2020). <https://doi.org/10.1007/978-3-030-63955-6>
8. Xiong, X., Peng, X., Cao, X.: Research on Oracle database heterogeneous resource integration method based on improved ORM. *Electron. Des. Eng.* **28**(21), 38–41+46 (2020)
9. Zhi, H., Tao, L., Yao, S., et al.: Research on synchronization strategy of small heterogeneous database based on Json. *J. Meteorol. Res. Appl.* **41**(1), 48–53 (2020)
10. Xiao, G.: Research and implementation of heterogeneous database update synchronization. *Software Guide* **18**(10), 182–185 (2019)
11. Liu, S., Liu, X., Wang, S., Muhammad, K.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021). <https://doi.org/10.1007/s00521-020-05021-3>
12. Xiong, H., Xu, D.: JSON based heterogeneous database integration model. *Digital Technol. Appl.* **38**(10), 33–35 (2020)
13. Li, R., Ren, Z., Huang, G., et al.: Design and implementation of heterogeneous architecture for database query acceleration. *Comput. Eng. Sci.* **42**(12), 2169–2178 (2020)



# Research on Homomorphic Retrieval Method of Private Database Secrets in Multi-server Environment

Fu-lian Zhong<sup>1(✉)</sup> and Jin-hua Liu<sup>2</sup>

<sup>1</sup> School of Mathematics and Computer Science,  
Xinyu University, Xinyu 338031, China  
zhongfulian682@163.com

<sup>2</sup> Department of Professional and Continuing Education, Xinyu University,  
Xinyu 338031, China

**Abstract.** The traditional homomorphism retrieval method of privacy database is very complicated. In order to reduce the running time, this paper designs a homomorphic secret retrieval method for private database in multi-server environment. After the establishment of the secret homomorphism vector model, the semantic classification of the secret homomorphism ciphertext is carried out. Then, according to the characteristics of the neighborhood structure, the mapping interval is divided, and the HASH function is used to perform operations in the mapping interval. This process can reduce the computational complexity of the secret homomorphic ciphertext. Finally, a secret homomorphic retrieval model is established and an optimized retrieval algorithm is designed. Design experiments and compare the three conventional retrieval methods. According to the experimental data, in different mapping intervals, the average retrieval time of this method is 14.32 s, while the average retrieval time of the three control groups is 23.74 s, 29.03 s, At 20.92 s, the retrieval time of this method is shorter than that of the conventional method, which makes the homomorphic retrieval method of private databases more concise.

**Keyword:** Multi-server environment · Privacy database · Secret homomorphism · Data retrieval

## 1 Introduction

Database system is the core component of computer information system, and its security problem is an important aspect to be studied in the field of information security. For some important or sensitive data, users are eager to store and transmit in the form of ciphertext, and can operate on the database information without revealing the content. The database encryption method can be applied to different environments, but there is a common problem that the formed ciphertext database cannot be operated. In other words, for a ciphertext database, if you want to perform mathematical

operations such as statistics, averaging, and summation on certain fields, you must first decrypt these fields, then perform mathematical operations on the plaintext, and then encrypt [1]. In this way, firstly, the space-time overhead is increased; secondly, in actual applications, for some important or sensitive data, it cannot meet the needs of users to operate on them without letting users know the information in them. If you can perform mathematical operations and regular database operations on the ciphertext database, you can obviously solve the above problems, and can greatly reduce the time and space overhead required for encryption/decryption, and improve the operating efficiency of the database. Secret homomorphism technology is an effective method to solve the above problems.

A homomorphism retrieval method for database secret is designed in Reference [2]. This technology uses a specific algorithm to segment the data first, and obtains its index number. The index number represents the data of the corresponding data segment. Although this method can greatly reduce the scope of the search, it can only improve the accuracy of the search because of the increase of the search steps, but cannot reduce the search time. Reference [3] uses the interaction between the server and the client to reduce the retrieval time. However, in the retrieval process of this method, the values that have been recorded with the same attributes are always mapped to the same set. After rigorous statistical processing, they are vulnerable to attacks. Reference [4] proposed a method of using the divide-and-conquer principle to build indexes on non-homomorphic ciphertexts, which realized fast data retrieval. However, such a retrieval method is very clear to the opponent, so it can quickly track and locate the database access process, and it is easy to expose the partial order relationship of the data. This method is difficult to deal with the dynamic analysis of the adversary, which makes it difficult to guarantee the security of the database.

In order to solve the shortcomings of the above methods, a new homomorphism retrieval method for private database in multi-server environment is designed in this paper. The design idea of the new method is as follows:

- (1) Based on the establishment of the secret homomorphic vector model, the semantic classification of the secret homomorphic ciphertext is carried out, so as to effectively reduce the search scope.
- (2) Divide the mapping interval according to the features of the ciphertext neighborhood structure, and then use the HASH function to perform operations within the mapping interval to reduce the computational complexity of the secret homomorphic ciphertext.
- (3) Establish the secret homomorphic retrieval model and design the retrieval algorithm to optimize the homomorphic retrieval process in the multi-server environment.

## 2 Design of a Homomorphic Retrieval Method for Private Database Secrets Based on a Multi-server Environment

### 2.1 Establish a Secret Homomorphic Vector Model

Firstly, the key parameters in the secret homomorphism of the private database are preprocessed, and a vector composed of multiple key parameters is used to represent the vector model of the secret homomorphism of the database to improve the accuracy of retrieval. On this basis, the secret homomorphism in the database is classified into ciphertext semantics, and the ciphertext semantic feature vector is extracted. At this time, a vector model should be established first, so that the vector can represent the secret homomorphism in this feature space, and all the ciphertext semantic sentences related to the secret homomorphism in the database are extracted to establish this vector model. Each feature in the vector model has its feature value, that is, the weight of the feature. Each ciphertext semantic keyword is one of the characteristic dimensions. Assuming there are  $x$  ciphertext semantic keywords in the database, the characteristic dimension in the database is also  $x$ . For a secret homomorphism  $d_j$ , the frequency of the ciphertext semantic keyword in the database can be used to calculate the weight of the ciphertext semantic keyword, and the similarity can be used as its continuous measurement index. The formula for calculating the frequency of occurrence of ciphertext semantic keywords is as follows:

$$TF : f_{ij} = \frac{freq_{ij}}{Maxfreq_j} \quad (1)$$

Among them,  $TF : f_{ij}$  represents the frequency of ciphertext semantic keywords,  $freq_{ij}$  represents the number of times the ciphertext semantic keywords appear in the secret homomorphism, and  $Maxfreq_j$  is the number of times the ciphertext semantic keywords appear most frequently in the secret homomorphism [5]. This formula is mainly used to calculate the relative frequency of a ciphertext semantic keyword in a secret homomorphism.

The ciphertext semantic keywords appearing in the secret homomorphism are converted into a vector in the order of frequency, the contents of the secret homomorphism are extracted, attributes and attribute values are extracted, and the ciphertext semantic vector describing the contents of the secret homomorphism is described. Let  $E_a$  and  $E_b$  represent the encryption and decryption steps of the plaintext, then the elements in the plaintext data space are a finite set that can become a secret homomorphic space with  $(E_{a1}, E_{a2}, E_{a3}, \dots, E_{an})$  as the main body [6]. In order to improve the efficiency of secret homomorphic retrieval, classify the secret homomorphic ciphertext semantics. Based on the above extraction of secret homomorphic ciphertext keywords, classify the relevant ciphertext semantic vectors and describe the hierarchical concept tree As shown in Fig. 1.

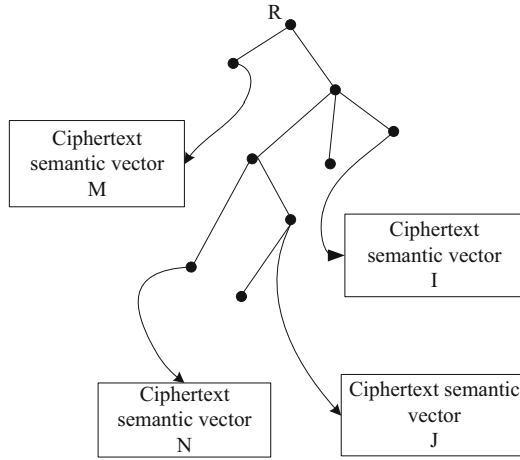


Fig. 1. Semantic classification of secret homomorphic ciphertext

As shown in Fig. 1, R is the source of the hierarchical concept tree. When the retrieval method in this paper classifies the ciphertext semantic vector, one source can be used as the origin of the hierarchical concept tree to derive other ciphertext semantic vectors, and finally deduced Multiple different ciphertext semantic vector branches provide a theoretical basis for the secret homomorphic ciphertext semantic classification index.

### 2.2 Refine Search Scope Based on Ciphertext Semantic Similarity

Each secret homomorphism has different ciphertext semantic characteristics, so when calculating the similarity of secret homomorphism, the problem of conceptual attributes needs to be considered first. Suppose concept A has an instance M. At this time, the instance M can be represented as  $M = A[P]$ , where  $P = (P_1, P_2, \dots, P_m)$  is the same as in the instance  $N = A[Q]$ ,  $Q = (Q_1, Q_2, \dots, Q_n)$ . At this time, the similarity of instance N and M can be calculated. First, the attribute vector of instance N and M is a common attribute vector through the above method, and then the similarity of the secret homomorphism is calculated according to the attribute value, and the attributes of the two instances are compared. Value and similarity, the obtained formula is shown below.

$$\begin{aligned}
 Sim_p(P, Q) &= Sim_p(P', Q') \\
 &= \sum_{i=1}^r \frac{\mu_1 + \gamma_1}{2} Sim_p(p', q')
 \end{aligned}
 \tag{2}$$

Among them,  $\mu_1$  and  $\gamma_1$  are the weight coefficients of attributes  $p'$  and  $q'$  in each vector, which is a preset parameter, usually the statistical value obtained after preprocessing the secret homomorphism, and the secret homomorphism is determined by this

statistical value. The final weight value of the state, the value range is [0.1]. Use deep learning to explore the search range of ciphertext semantic similarity. Generally, a mathematical model is established through a convolutional neural network, and the time sequence information of the text data is considered. The data information output at the current moment is regarded as the data information output at the previous moment, and input at the current moment. The superposition of data information [7, 8]. And build the entire convolutional neural network into a three-in-one model structure including the input layer, hidden layer, and output layer. Through the training of the neural network, a more detailed search range of ciphertext semantic similarity is obtained, and the text data information is processed Classification processing, the specific structure of this classification model is shown in the figure below.

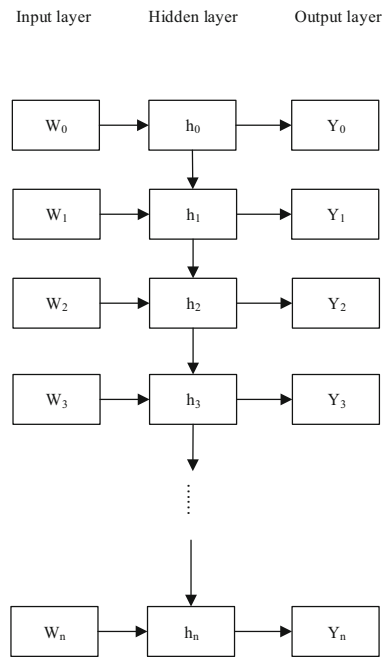


Fig. 2. Deep learning model structure

As shown in Fig. 1, assuming that the data text of the input layer is  $w = (w_0, w_1, w_2, w_3, \dots, w_n)$ , the text data of the input layer is sequentially converted to the hidden layer, so that the data becomes  $h = (h_1, h_2, h_3, \dots, h_n)$ , and then through the trained neural network model, the hidden layer data is converted into the data text  $y$  of the output layer through calculation. The calculation process is as follows:

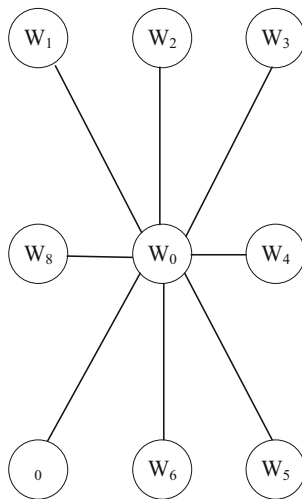
$$\begin{cases} h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \\ y_t = W_{hy}h_t + b_y \end{cases} \quad (3)$$



Among them,  $W_{xh}$  represents the function converted from the input layer to the hidden layer,  $W_{hh}$  represents the internal conversion function of the hidden layer,  $W_{hy}$  represents the function from the hidden layer to the output layer,  $b_h$  represents the deflection vector from the input layer to the hidden layer, and  $b_y$  represents the deflection vector from the input layer to the hidden layer. The deflection vector from the hidden layer to the output layer. The update mode of the input layer, hidden layer and output layer can be obtained by formula (3), and the state information of the input layer and output layer at the previous moment can be obtained by training the hidden layer, and then combined with the convolutional neural network to obtain the input at the current moment And output text information, and to achieve the purpose of reducing the search range of ciphertext semantic similarity [9]. Train the objective function through the convolutional neural network, and redefine the function that changes after passing through the hidden layer:

$$L = - \sum_{i=1}^T Y_i \log(y_i) \tag{4}$$

Among them,  $T$  represents the total amount of text information that needs to be classified,  $Y_i$  represents the product of the probability distribution of each category after the text information is classified and its predicted value, and  $y_i$  represents the probability distribution value of each category after the text information is classified. Then, the neighborhood structure of edge points and noise points in the search range of ciphertext semantic similarity is set as shown in Fig. 3.



**Fig. 3.** Neighborhood structure features

As shown in the figure above,  $W_0$  is the center of the entire image, and there are 8 neighborhoods around it. The connection between each neighborhood and the point  $W_0$  of the image center is defined as:

$$Set_t = \{(0, t)/1, 2, \dots, 8\} \quad (5)$$

In the formula,  $t$  is the connection from any point  $t$  to the center 0 point. Therefore, it can be considered that when the center 0 point is at the edge of the image, any  $Set_t$  is around the center 0 point, that is, the edge of the neighborhood structure [10]. The average gradient direction of the center 0 point and any point  $t$  in its surrounding neighborhood is defined as  $Ang_t$ . The noise point of  $Ang_t$  is usually large. Therefore, it is necessary to suppress noise and construct a neighborhood structure through a deep learning model to reduce the search range.

### 2.3 Divide the Mapping Interval

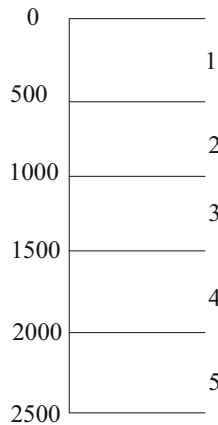
After narrowing the scope of database secret homomorphism retrieval, the method designed in this paper can basically obtain the similarity measurement method of ciphertext semantics, and accurately reflect the similarity between two ciphertexts of the same type. It directly combines the two ciphertexts. The text is quantified as a numerical value to indicate the degree of similarity for better retrieval. At this time, the similarity of the two ciphertexts can be expressed by the method of mapping intervals. The greater the distance, the greater the difference between the two ciphertexts, and the greater the distance, the smaller the difference between the two ciphertexts. The distance formula used is different, which will directly affect the accuracy of the retrieval algorithm. This paper uses the chi-square distance method to measure the similarity between two ciphertexts. The formula for calculating the chi-square distance is as follows:

$$D(a, b) = \sum_t \frac{(a_t - b_t)^2}{(a_t + b_t)} \quad (6)$$

Among them,  $a_t$  represents the sample image  $A$ ,  $b_t$  represents the image  $B$  to be retrieved in the image database;  $D(a, b)$  represents the chi-square distance between the sample image  $A$  and the image  $B$  to be retrieved [11, 12]. In this paper, the similarity measurement algorithm of two ciphertexts is constructed by the chi-square distance. When the chi-square distance is 0, it means that the two ciphertexts are exactly the same. After the distance of the mapping interval is obtained, all the necessary data contained in it can be obtained through the secret homomorphism search that has reduced the search space. To perform a secret homomorphic search on the ciphertext database, the ciphertext data must be divided into intervals. According to the retrieval conditions, several divided contents are returned, and the returned contents must contain all the required data. According to the obtained index value, we divide it according to the specified rules. The division must meet the following conditions: First, all division intervals must cover all data. Second, the division intervals cannot overlap. Third, there is and can only correspond to one interval after the specific value mapping. Assuming that the divided intervals are continuous, the maximum and minimum values of each interval can be found. Use the traditional equation to define the partition function, which is described as follows:

$$\Psi(W_i, W_j) = \{\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_n\} \tag{7}$$

In the formula,  $W_i$  represents the traditional divided interval that has been abandoned;  $W_j$  represents the divided interval currently in use;  $\{\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_n\}$  represents the sub-interval segmentation of an interval. According to the above division function, select an appropriate division method, then the data to be retrieved can be divided into  $n$  parts. However, for a large database, there is a lot of data in each divided segment. After the secret homomorphic retrieval, the server returns the result of the secret homomorphic retrieval to the client for accurate retrieval, and the scope of the retrieval is also very large. This paper proposes a sub-division method. On the basis of the traditional division function, each division segment is continued to be sub-divided. Such a division method has very important significance for the accuracy of future secret homomorphic retrieval. Different fields may require different division rules and functions, such as division by breadth, division by depth, etc. [13]. There are many ways to divide, and the selection requirements must meet the three requirements mentioned above. After the division is successful, in order to facilitate the secret homomorphic retrieval, the divided interval is replaced with a specific value or similar value that can be marked. Here, a collision-free HASH function is needed, and the HASH function maps the divided interval to a specific mark value. At the same time, the HASH function can be used to mark the divided intervals one by one. For example, at this time, the upper bound of the interval and the divisor of 40 can be selected as the interval marker. The specific interval markers are shown in Fig. 4.



**Fig. 4.** Secret homomorphic segmentation interval division

It can be seen that the data processing is divided into two steps. The first step is to divide the overall data into equal parts. In the second step, the sub-division interval is refined according to the division function of the first step. The marking of the interval does not necessarily need to be in an increasing order, as long as it is satisfied that the

HASH is collision-free within the specified range, the HASH function can be used to perform the marking operation.

### 2.4 Establish a Secret Homomorphic Retrieval Model

The secret homomorphic search is performed on the basis of the newly established ciphertext index. Figure 3, 4 shows the specific implementation process. There are three main steps in secret homomorphic retrieval, condition mapping, retrieval sentence conversion, and retrieval. The main purpose of secret homomorphic search is to reduce the search scope of precise search, so that the client search time can be greatly shortened [14]. The idea of secret homomorphic retrieval introduced in this article is.

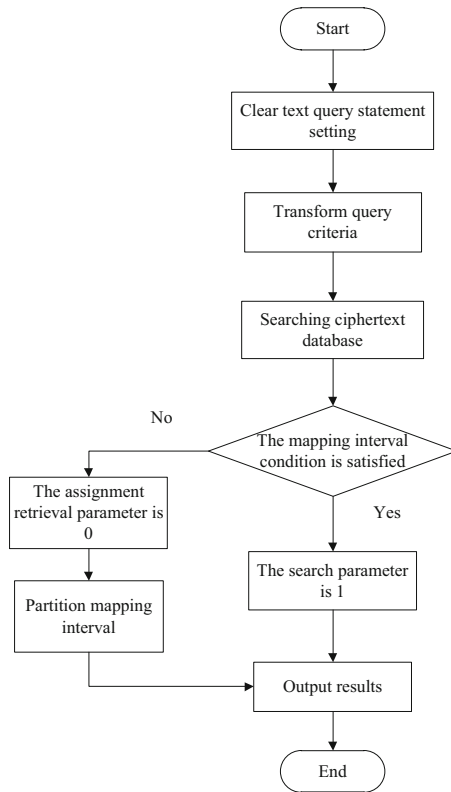


Fig. 5. Algorithm structure design

As shown in Fig. 5, when inserting data into the database, the data is processed first. According to the characteristics of the data itself, the data is grouped for storage. And to select a unique value to mark the group. Transform the search conditions and associate them with the grouping mark, so that the plaintext and the ciphertext are reconnected. When storing ciphertext data, add a flag option after each record. The

initial value of the flag option is 0. In the second step, it is judged whether all the divided intervals meet the conditions. If all the conditions are met, the flag flag is set to 1. Otherwise, compare the search conditions with each partition according to the search conditions and the divided partitions, exclude the partitions that do not meet the conditions, and return the partition data that contains the correct search object. For partitions that partially meet the conditions, the search conditions are compared with their sub-partition marks. Return the numerical region corresponding to the sub-division mark as the result of the secret homomorphic search. In this process, dividing the interval is a key factor affecting the accuracy and time consumption of secret homomorphic retrieval. The finer the division interval, the higher the accuracy of the retrieval.

### 3 Experimental Study

#### 3.1 Experiment Preparation

Compared with translating ciphertext into plaintext form and then performing mathematical calculations, secret homomorphism has a great advantage of simplicity. Especially in the field of ciphertext retrieval, traditional retrieval methods are very complicated, but secret homomorphism can be fast and efficient. Get the information that needs to be retrieved and display it to the customer. This experiment is mainly used to test whether the database secret homomorphism in a multi-server environment can be accurately searched while improving retrieval efficiency, and the answer can be obtained through the comparison of retrieval time. The server in this experiment is divided into a master and a client, and its overall architecture is shown in Fig. 6.

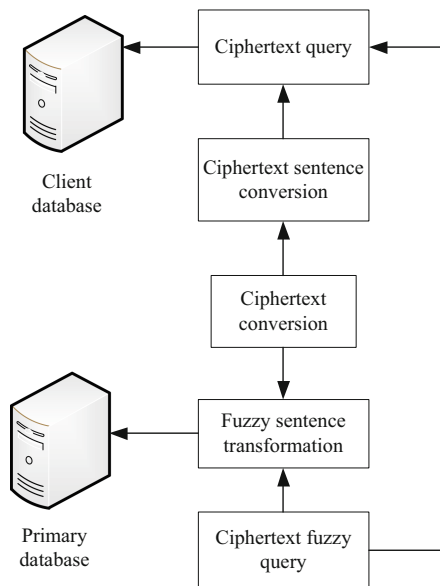


Fig. 6. Server architecture

As shown in Fig. 6, the main frame of the server architecture design is divided into two parts, the client part and the master part. The main task of the server part is to conduct secret homomorphic search, narrow the search scope of precise search, and reduce the search time of the client. The client mainly performs precise search and decrypts the information returned by the server. Among them, the flow of search sentences and non-search sentences is different [15]. Since the retrieval statement has a specific return result, it needs to be temporarily stored on the client side to facilitate accurate retrieval. In Fig. 6, the exact search sentence and the secret homomorphic search sentence are different secret homomorphic search sentences, and the search conditions need to be mapped to the ciphertext state according to the ciphertext index to perform the secret homomorphic search. Since the ciphertext index is excluded from the results returned by the secret homomorphic search sentence, the precise search sentence only homomorphically encrypts the search condition to facilitate accurate search. This experiment is mainly divided into two aspects. First, the performance of the retrieval method is tested, and the retrieval performance of the master server and the client server are tested respectively. Second, it takes time to test the ciphertext retrieval function of the secret homomorphism technology. According to the test needs, this experiment will give the time consumption of the plaintext to judge. Compare the conventional retrieval method with the retrieval method in the text, and analyze whether the method in this paper has realized the optimization of convenience.

### 3.2 Classification Performance Test

The test analysis of homomorphic technology is mainly to detect the ciphertext expansion after the homomorphic encryption processing. In this process, it is necessary to detect the ciphertext expansion of the client database and the master database separately. When calculating the expansion of the data, it is necessary to determine the main factors of its expansion according to the definition of modulo operation in the previous article, and combine the main factor  $n$  to obtain the passive factor  $P$ . On the premise that  $P$  is a large prime number, the effect of data expansion is very significant. The database table design of this experiment is shown in Table 1.

**Table 1.** Database ciphertext expansion test table

Field	Symbol	Character type	Is it empty?
Plaintext 1	N1	Float	No
Plaintext 2	N2	Float	No
Plaintext 1 + Plaintext 2	$N1 + N2$	Float	No
Plaintext 1 $\times$ Plaintext 2	$N1 \times N2$	Float	No
Ciphertext 1	M1	Float	No
Ciphertext 2	M2	Float	No
Ciphertext 1 + ciphertext 2	$M1 + M2$	Float	No
Ciphertext 1 $\times$ ciphertext 2	$M1 \times M2$	Float	No
Plaintext 1 to ciphertext 1	$N1 \rightarrow M1$	Float	No
Plaintext 2 Convert ciphertext 2	$N2 \rightarrow M2$	Float	No

Combining the storage comparison of plaintext data and ciphertext data, the ciphertext storage data of the main-end database and the client-side database can be obtained as shown in Table 2 and Table 3. Among them, the retrieval method of this article is set as the experimental group, and the conventional retrieval method is set as the control group 1-the control group 3.

**Table 2.** Main-end database ciphertext storage

	Test group	Control group 1	Control group 2	Control group 3
N1	4.9	-5.6	7.2	9.3
N2	8.6	12.3	-3.5	10.7
N1 + N2	16.7	13.2	12.9	-10.1
N1 × N2	52.253641	24.639745	-12.369742	39.621452
M1	25324.202	12536.745	16328.963	-45863.142
M2	25639.142	12574.369	-14256.745	41253.963
M1 + M2	425639.112	458963.745	-125798.312	256364.756
M1 × M2	225896343.526	274896512.369	-147963428.258	546328961.248
N1 → M1	256342.96	412563.75	-142586.48	127463.52
N2 → M2	563214.85	-178563.42	145286.42	463285.96

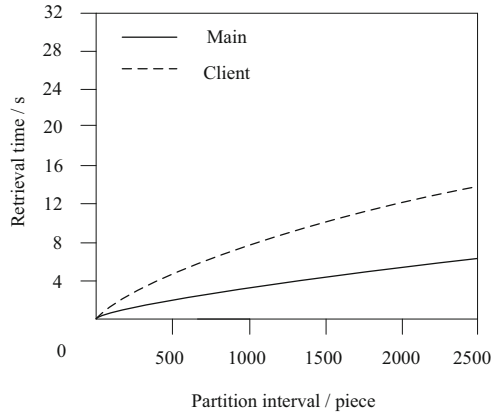
**Table 3.** Client database ciphertext storage

	Test group	Control group 1	Control group 2	Control group 3
N1	2.3	-3.5	7.5	8.3
N2	8.4	12.1	-2.3	9.7
N1 + N2	12.6	15.8	16.1	-10.4
N1 × N2	56.32414	23.56915	-14.36982	34.12572
M1	12356.74	45698.25	53152.75	-14763.52
M2	45628.362	41526.760	-14863.285	47856.251
M1 + M2	425479.168	428568.775	-136947.307	275463.149
M1 × M2	456975246.185	852643257.125	-964531752.365	415862463.75
N1 → M1	258963.14	128476.25	-163725.69	158963.12
N2 → M2	746321.85	-145286.25	756395.12	415963.14

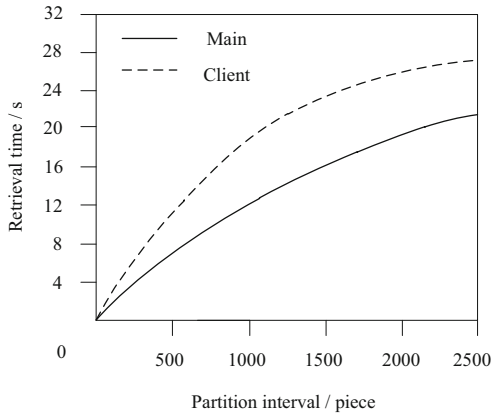
From Table 2 and Table 3, it can be seen that the ciphertext data in the main-end database and the client-side database are expanded very obviously, and the specific expansion multiple is related to the size of r. The storage size of the ciphertext of the experimental group is about three times that of the three control groups. To four times. And such ciphertext expansion coefficient proves that the secret homomorphic retrieval method of private database based on multi-server environment in this paper has good performance.

### 3.3 Actual Retrieval Time Test

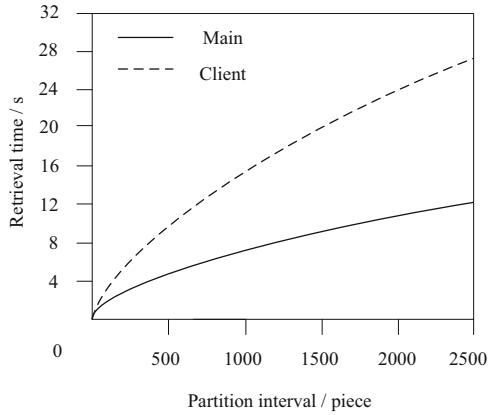
The biggest factor that affects the retrieval time is the size of the data ciphertext division interval, or the number of intervals. Figure 7 shows the time consumption



(a) test group



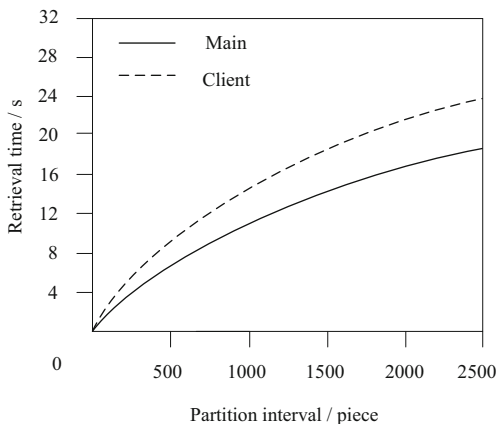
(b) Control group 1



(c) Control group 2

**Fig. 7.** Retrieval time test





(c) Control group 3

Fig. 7. (continued)

change diagram of the system in the search operation with join, as the division interval increases, and the number of sub-division intervals does not change.

In general, the total time consumed by the host and the client is shown in Table 4 under the premise that the division interval of the four retrieval methods is increasing.

Table 4. Total time consumption

	Test group	Control group 1	Control group 2	Control group 3
100	2.14	2.75	5.62	4.15
200	3.25	4.36	8.63	6.26
300	4.58	7.41	11.41	8.42
400	7.12	10.25	14.25	10.85
500	8.09	13.59	17.42	12.28
600	9.16	14.75	19.14	14.25
700	10.23	16.62	21.26	15.34
800	11.34	18.51	23.34	17.41
900	12.45	20.76	25.75	18.28
1000	13.17	22.18	26.17	19.05
1100	14.03	23.16	27.14	20.36
1200	14.84	24.28	28.45	21.42
1300	15.46	26.36	30.36	22.96
1400	15.75	27.15	32.75	23.42
1500	16.23	28.52	33.42	24.08
1600	16.94	29.36	34.42	25.14

(continued)

**Table 4.** (continued)

	Test group	Control group 1	Control group 2	Control group 3
1700	17.65	29.41	36.52	26.32
1800	18.12	30.85	37.61	26.85
1900	18.97	31.14	39.07	27.64
2000	19.56	32.15	40.12	28.05
2100	20.36	32.47	41.14	28.95
2200	21.04	33.52	42.15	29.42
2300	21.85	34.63	42.16	29.96
2400	22.52	35.46	43.32	30.53
2500	23.25	36.78	44.15	31.52

According to Fig. 7 and Table 4, it can be seen that the running time of the four retrieval methods is related to the division function. If the fields to be operated are divided into more intervals, the overall time spent is reduced. The finer the interval division, the finer the data retrieved by the secret homomorphic search. However, when the interval is divided to a certain extent, the time consumption tends to stabilize and approach a stable value. Explain that although the division of intervals has a great influence on the system time consumption, within a certain range, the consumption time will decrease as the division of intervals increases. Beyond a certain time limit, it remains stable. Among them, the average retrieval time of the experimental group was 14.32 s, while the average retrieval time of the three control groups were 23.74 s, 29.03 s, and 20.92 s. According to the above experimental results, the proposed homomorphic secret retrieval method of private database has a shorter running time than the three traditional methods, which proves that its retrieval process is simpler.

## 4 Conclusion

The privacy of database is always the core problem of information security. However, the process of converting plaintext to ciphertext also makes the computation cumbersome. After a lot of practice, the secret homomorphism of database is the key to solve the problem of convenient ciphertext calculation of database. To solve this problem, this paper designs an effective homomorphism retrieval method for privacy database. In the experiment part, the conversion and computation time of ciphertext data are tested. Compared with the three traditional methods, the proposed method consumes much less time, so it can be seen that the proposed method can calculate ciphertext data more quickly and conveniently.

In the following research, the method in this paper will be further optimized to improve its retrieval accuracy.

## References

1. Zhou, N., Zhang, M.-q., Liu, M.-m.: Reversible data hiding algorithm in homomorphic encrypted image based on secret sharing. *Sci. Technol. Eng.* **20**(19), 7780–7786 (2020)
2. Tan, Y., Lu, L., Wang, J.: Ciphertext-policy attribute encryption scheme based on homomorphic encryption. *Comput. Eng. Appl.* **55**(19), 115–120+127 (2019)
3. Li, S., Jing, Z.: Analysis and development strategy of cross-language retrieval function for “the belt and road” multilingual shared database. *Library Inf. Ser.* **65**(3), 20–27 (2021)
4. Liu, Y.-j.: Design of rapid retrieval system of archives information database based on MapReduce. *Electron. Des. Eng.* **28**(13), 45–49 (2020)
5. Wang, N., Zheng, K., Fu, J., et al.: Method of ciphertext retrieval in mobile edge computing based on block segmentation. *J. Commun.* **41**(7), 95–102 (2020)
6. Xiang, C.H.E.N., Heng, H.E., Peng, L.I., et al.: Ciphertext image retrieval scheme based on target detection in cloud environment. *Comput. Eng. Appl.* **56**(11), 75–82 (2020)
7. Yang, X., Chen, G., Li, T., et al.: Multi-user ciphertext retrieval scheme based on certificateless cryptosystem. *Comput. Eng.* **46**(9), 129–135 (2020)
8. Liu, J., Zheng, X., Zheng, D., et al.: Secure attribute based encryption enabled cloud storage system with ciphertext search. *Netinfo Secur.* (7), 50–58 (2019)
9. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>
10. Zhongyuan, Q.I.N., Yin, H.A.N., Xuejin, Z.H.U.: Research on ciphertext full-text retrieval of cloud storage based on improved DGHV algorithm. *Netinfo Secur.* **1**, 8–15 (2019)
11. Liu, S., Liu, X., Wang, S., Muhammad, K.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
12. Dai, H., Yang, G., Min, Z.: Multi-keyword parallel ciphertext retrieval scheme in distributed environment. *J. Comput. Appl.* **39**(10), 2948–2954 (2019)
13. Zhang, Y., Liu, X., Lang, X., et al.: Multi-server key aggregation searchable encryption scheme in cloud environment. *J. Electron. Inf. Technol.* **41**(3), 674–679 (2019)
14. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
15. Shu, Q., Wang, S., Han, L.: Analysis and improvement of a three-factor authentication protocol in the multi-server environment. *J. Hangzhou Normal Univ. (Nat. Sci. Ed.)* **20**(1), 91–94+112 (2021)



# Low-Frequency Noise Characteristic Extraction Method of Electronic Components Based on Data Mining

Xiao-jing Qi<sup>(✉)</sup>

Chongqing Telecommunication Polytechnic College, Chongqing 402247, China  
qixiaojing2021@163.com

**Abstract.** Aiming at the low accuracy of traditional extraction methods, a method based on data mining is proposed to extract low-frequency noise features of electronic components. Firstly, data mining is carried out. Based on this, the noise modulation model of electronic components is established. Combined with filtering, the low-frequency noise feature extraction method of electronic components is optimized, and the experimental analysis is carried out. The experimental results show that this method can effectively improve the accuracy of low frequency noise characteristics of electronic components, and has a certain practicality.

**Keywords:** Electronic components · Low-frequency noise · Characteristics · Extraction · Data mining

## 1 Introduction

Low frequency electrical noise of electronic devices is the manifestation of micro motion of electronic carrier. Its accurate measurement can provide the basis for noise characteristics, generation mechanism and analysis application. Noise measurement can effectively verify the basic theories related to carrier transport in electronic components, and also promote the exploration of the physical sources of noise phenomenon. The application of electric noise is one of the goals of noise mechanism research and noise testing technology research. Based on the research of low-frequency electrical noise, the correlation research of electronic component defects, stress damage, process level, quality and reliability is put forward, and applied to the optimization design of components, process control, quality evaluation and reliability screening.

Accurately measure the performance of carrier micro motion in low frequency noise of electronic devices, which lays a foundation for the study of noise characteristics, generation mechanism, analysis and application. Noise is a common phenomenon in nature. People's understanding of noise begins with the most intuitive sense of hearing and sight. The response of the auditory system and the visual system to the noise of the electronic system are “noise” and “noise” respectively [1]. The noise of electronic components restricts the noise level, detection sensitivity, false trigger rate, fidelity and resolution of the whole system. With the deepening of research, people gradually realize the dialectical relationship between noise and signal, that is,

noise contains blood signal. Therefore, the connotation of noise research has been gradually expanded, and the extraction and detection of noise have also begun to receive attention. The internal noise of electronic components, especially the low frequency noise, is very sensitive to the difference of bystanders or the defects caused by different materials, structures and processes of hardware components. Therefore, the low-frequency noise of electronic components can also be applied to the study of device quality and reliability characterization [2, 3]. The research on noise extraction and measurement of electronic components needs the efforts and wisdom of scholars in various fields. On the one hand, the deep understanding and explanation of the essential laws of physical phenomena in the theoretical study of noise is the basis of noise testing, analysis and application research. On the other hand, the study of noise testing and analysis is the verification and supplement of theoretical research, and also the power of noise extraction and testing technology progress and application transformation. For this reason, this paper proposes a method of low frequency noise feature extraction of electronic components based on data mining. The innovation lies in the use of autoregressive model to extract the deterministic components and obtain the autoregressive model of low frequency noise. By linear transformation of the autocorrelation function of low frequency noise, the weight coefficient of linear prediction can be obtained. Experimental results show that this method can effectively improve the accuracy of low frequency noise characteristics, and has certain practicability.

## 2 Data Mining

Assuming that  $\{X_n\}$ ,  $n = 1, \dots, N$  is the noise point of an electronic component, which is located in a phase space, and  $C(\varepsilon)$  is the correlation function. Then the proportion of points whose distance is less than a certain value can be expressed as:

$$C(\varepsilon) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N \theta(\varepsilon - \|X_i - X_j\|) \quad (1)$$

Among them,  $\|X_i - X_j\|$  is the distance represented by the norm, and  $\theta$  represents the Heaviside step function. When  $x$  takes different values,  $\theta$  will take two different numbers. For the point pair  $(X_i, X_j)$  with a distance less than a certain value, it is recorded as the correlation dimension. When the distances of  $N$  meet certain requirements,  $C(\varepsilon)$  obeys the exponential law and needs to meet:

$$D = \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow 0} d(N, \varepsilon) \quad (2)$$

The above formula  $D$  is the correlation dimension. Envelope demodulation is a commonly used feature extraction method, which mainly includes two parts: the solution of analytical electronic component signals and the Fourier transform. Analyzing the solution of electronic component signals is a key technology [5], and the solution method is shown in formula (4):

$$x_H \sqrt{x(n, f, \Delta f)^2 + (\text{Hilbert}(x(n, f, \Delta f)))^2} \quad (3)$$

In the formula,  $x_H$  is the envelope signal of the frequency band signal, and  $\text{Hilbert}(\cdot)$  is the Hilbert solution function.

On this basis, select a reference point and make a neighborhood, find the trajectory distance from the neighborhood point to the reference point [6], and take the logarithm of the average distance to obtain the relevant data of the electronic components. The calculation formula is:

$$S(\nabla n) = \frac{1}{N} \sum_{n0=1}^N \ln \left( \frac{1}{|u(\beta_{n0})|} \sum_{\beta_n \in u(\beta_{n0})} |s_{n0} + \nabla n - s_n + \nabla n| \right) \quad (4)$$

In the formula,  $\beta_{n0}$  and  $u(\beta_{n0})$  are the embedded vector and neighborhood respectively,  $|u(\beta_{n0})|$  is the relevant data volume of electronic components,  $\beta_n$  is the salient feature point of electronic components,  $s_{n0}$  is the last salient element of  $\beta_{n0}$ , and  $s_{n0} + \nabla n$  represents the time range beyond.

### 3 Extraction Method of Low-Frequency Noise Characteristics of Electronic Components Based on Data Mining

#### 3.1 Noise Modulation Model of Electronic Components

The noise generated in the operation of electronic components contains obvious modulation signals, which are mainly caused by the periodic excitation force of electronic components. In this regard [7], the establishment of a single-component modulation model of electronic component noise is shown in the following formula.

$$x_{MA}(t) = A_m \cos(2\pi f_{ci}t) \quad (5)$$

In the formula,  $x_{MA}$  represents the noise modulation signal of electronic components,  $A_m$  is the amplitude of the noise modulation signal,  $f_m$  is the characteristic frequency of the modulation signal,  $f_{ci}$  is the frequency of the carrier signal, and  $N$  is the total number of carrier signals.

As a result, a noise modulation signal appears in the emitted noise of electronic components, and the established amplitude modulation-frequency modulation modulation signal of the low-frequency noise signal of electronic components is shown in formula (7). The characteristic frequency of the modulation signal can be solved by the derivative of the corresponding information [8], and its instantaneous modulation frequency is shown in formula (8).

$$x_{MAf}(t) = A_{mf} \cos(\theta_{mf}(t)) \sum_i^n \cos(2\pi f_i t) \quad (6)$$

In the formula,  $x_{MAf}(t)$  represents the low-frequency noise signal,  $A_{mf}$  is the amplitude of the low-frequency noise signal, and  $\theta_{mf}(t)$  is the change of the angle of the low-frequency noise modulation signal over time.

$$f_{mf}(t) = \frac{d\theta_{mf}(t)}{dt} \quad (7)$$

In the formula,  $f_{mf}(t)$  represents the instantaneous frequency of the low-frequency noise signal.

The autoregressive model is used to extract the deterministic components. The autoregressive model of low-frequency noise can be expressed by the following formula:

$$x_{AR} = - \sum_{i=1}^p q_i x(n-i) \quad (8)$$

In the formula,  $x_{AR}$  represents the deterministic signal component obtained by linear prediction of low-frequency noise,  $q(i)$  represents the weight coefficient, and  $p$  represents the order of the AR model. The weight coefficient  $a(k)$  of linear prediction can be obtained by linear transformation of the low-frequency noise autocorrelation function  $r_{AA}(i)$ . The weight coefficient  $q(k)$  can be solved using Yule-Walker equations, and the noise modulation model of electronic components is obtained as shown below.

$$r_{xx}(i) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n-i), 0 \leq i \leq p-1 \quad (9)$$

### 3.2 Normalized Processing of Low-Frequency Noise Performance of Electronic Components

When the low-frequency noise intensity  $D = 0$ , there is a critical value  $A_\tau = \sqrt{4a^3/27b}$ . When  $A_0$  is greater than, the low-frequency noise balance of the entire electronic components will no longer be maintained. For this reason, for simplicity, let  $b = 1$ , and the bistable potential function is  $U(x) = -ax/2 + x^4/4$ . For small parameter signals (amplitude, frequency and noise intensity  $\ll 1$ ), as the time of occurrence of resonance of electronic components increases [9], the signal-to-noise ratio of the response of electronic components reaches the maximum. At this time, the output power spectrum  $S(f)$  consists of two parts of the power spectrum, namely:

$$S(f) = S_1(f) + S_2(f) \tag{10}$$

In the formula,  $S_1(f)$  represents the power spectrum caused by the input low-frequency noise periodic signal, which is equal to the frequency spectrum of the input signal;  $S_2(f)$  represents the power spectrum caused by the low-frequency noise, showing the form of Lorentz distribution [10]. Among them, the Lorentz distribution characteristic curve is shown in Fig. 1, and the respective expressions are as follows:

$$S_1(f) = \frac{2a^4A^2 \exp(-a/2D)/(\pi^2D^2)}{(2a^2 \exp(-a/2D)/\pi^2) + (2\pi f_0^2)} \times \delta(f_0 - f) \tag{11}$$

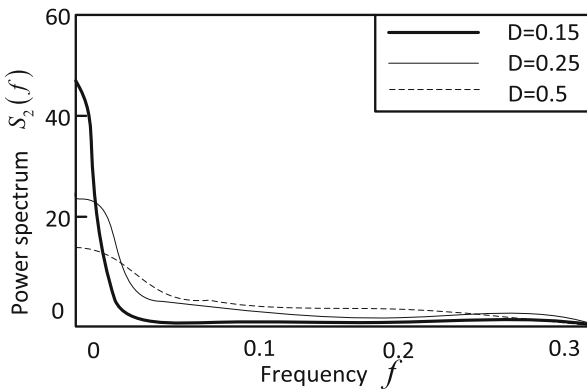


Fig. 1. Characteristic curve of Lorentz distribution

When the value of low-frequency noise intensity  $D$  is different, the corresponding noise output power spectrum  $S_2(f)$  varies with frequency. It can be seen that within a certain range, the greater the noise intensity, the smoother the Lorentz distribution characteristic curve; the smaller the noise intensity, the curve The steeper [11]. The output power of noise has the characteristic of Lorentz distribution.

The method of extracting the low-frequency characteristics of electronic components by entropy is usually to estimate the entropy by  $C_2(m, \varepsilon)$ . The main principle is as follows: Let  $\hat{C}_q(\varepsilon)$  represent the generalized  $C_2(m, \varepsilon)$  estimate, and the expression between  $\hat{C}_q(\varepsilon)$  and entropy is as follows:

$$C_q(m, \varepsilon) \propto \varepsilon^{(1-q)D_q} e^{(q-1)H_q(m)} \tag{12}$$

$D_q$  is the  $q$ -order generalized dimension, and  $H_q(m)$  is the generalized sum of entropy. If you want to divide the low-frequency noise performance of electronic components into countless parts, the parameter  $\varepsilon$  needs to meet the  $\varepsilon \rightarrow 0$  condition, but at the same time it will bring about some problems, that is, the number of points in the field is too small, resulting in obvious fluctuations, and it is difficult to apply it in practice [12]. In



order to have practical application value, the following conditions must be met: Entropy is independent of parameter  $\varepsilon$ . The usual practice is to determine the appropriate scale range, draw the  $C_2(m, \varepsilon)$  local slope  $d(m, \varepsilon) \sim \varepsilon$  curve, and observe the visible platform of each curve. If there is a visible platform, the size of  $D_p$  does not change, when  $\varepsilon$  selects a different value, the value of  $\varepsilon^{D_q}$  is always the same, then  $h_q$  can be obtained by the following formula:

$$h_q(m, \varepsilon) = H_q(m + 1, \varepsilon) = \ln \frac{C_q(m, \varepsilon)}{C_q(m + 1, \varepsilon)} \tag{13}$$

According to the characteristics of noise modulation of electronic components, the frequency of carrier noise is much greater than that of modulation noise, so the first wavelet packet decomposition frequency band is not included in noise demodulation [13]. The energy coefficient of the noise decomposition frequency band can be expressed by the proportion of the energy of each frequency band in the total energy, as shown in the following formula:

$$\eta_i^n = \frac{1}{2^n} \cdot 100\% \tag{14}$$

$$\sum_{i=2} E_i^n$$

In the formula,  $\eta_i^n$  represents the energy coefficient of the  $i$ -th frequency band of the  $n$ -th order.

According to the definition of the spectral correlation density function, the corresponding signal  $x(t)$ , its spectral correlation density function is [14]:

$$S_x(\omega, f) = \frac{1}{T} X(f + \omega/2) X^*(f - \omega/2) \tag{15}$$

From Eq. (15), it can be known that the spectral correlation density function of signal  $x(t)$  represents the correlation of its spectrum on two different frequency components  $f + \omega/2$  and  $f - \omega/2$ . The weaker the correlation, the more characteristic the low-frequency noise performance at the cycle frequency is  $\omega$ . Significantly. In order to be able to quantitatively describe the strength of the low-frequency noise performance of electronic components [15, 16], the statistic  $\rho_x(\omega, f)$  is defined as the spectral coherence function of the low-frequency noise  $x(t)$ , which normalizes  $S_x(\omega, f)$ :

$$\rho_x(\omega, f) = \frac{S_x(\omega, f)}{\sqrt{S_x(0, f + \omega/2) S_x(0, f - \omega/2)}} \tag{16}$$

### 3.3 Low-Frequency Noise Performance Extraction of Electronic Components

In order to simplify the extraction process, assuming that the low-frequency noise of electronic components is dominated by white noise [17], it can be obtained from Eqs. (19) and (20):

$$\frac{dx}{dt} - U'(x) + s(t) + n(t) \quad (17)$$

In the formula,  $s(t)$  is the input signal,  $n(t) = \sqrt{2D}\xi(t)$  is the white noise,  $\xi(t)$  is the Gaussian white noise,  $n(t)$  satisfies  $\langle n(t) \rangle = 0$ ,  $\langle n(t)n(t-\tau) \rangle = 2D\sigma(\tau)$ ,  $D$  is the noise intensity,  $U(x)$  is the potential function of the electronic component, and  $x(t)$  is the output signal of the electronic component [18]. Then the potential function of the electronic component, the expression is as follows:

$$U(x) = -\frac{ax^2}{2} + \frac{bx^2}{4} \quad (18)$$

In the formula,  $a$  and  $b$  are parameters whose potential function of electronic components is greater than 0. The potential function has a minimum value  $U(x) = -a^2/4b$  at  $x = \pm\sqrt{a/b}$  and a maximum value  $Ux = 0$  at  $x = 0$ .

According to the linear prediction, the deterministic signal composition of the signal can be obtained, and then the residual signal formula composed of the low-frequency noise modulation signal component and the noise signal can be obtained as follows [19, 20]. Therefore, linear prediction based on AR model is an effective signal pre-whitening tool.

$$x_{AM}(t) + x_e(t) = x(t) + \sum_{k=1}^p a(k)x(t-k) \quad (19)$$

In the formula,  $x_{AM}(t)$  represents the modulation noise signal, and  $x_e(t)$  is the noise component in the monitoring signal.

The choice of modulation frequency band has an important influence on the demodulation effect of the final algorithm. When the carrier frequency band is modulated, the modulation frequency band is mostly the frequency band with higher energy. In this study [21], the modulation noise energy is used as the selected basic index. The calculation of frequency band energy is shown in the following formula.

$$E_i^n = \|wpc_i^n\|_2^2 \quad (20)$$

In the formula,  $\|\cdot\|_2$  is the solution function of 2 norm,  $E_i^n$  represents the energy of the  $i$ -th frequency band of the  $n$ th order, and  $wpc_i^n$  represents the signal of the  $i$ -th frequency band of the  $n$ th order [22].

By synergizing the low-frequency noise and intensity of electronic components, a simple model that can produce stochastic resonance is obtained:

$$\frac{dx}{dt} = ax - bx + A_0 \cos(2\pi f_0 t) + n(t) \tag{21}$$

In the case of no change, the potential function of the electronic component is constrained by the periodic signal  $s(t) = A_0 \cos(2\pi f_0 t)$ , and the modulated bistable potential function is obtained as:

$$U(x) = -\frac{1}{2}ax^2 + \frac{1}{4}bx^4 - A_0x(2\pi f_0 t) \tag{22}$$

The ratio of the low-frequency noise of electronic components to the average power of the unit noise spectrum at  $f = f_0$  is called the signal-to-noise ratio of electronic components, namely:

$$SNR_{out} = \frac{\int_0^\infty S_1(f)df}{S_2(f = f_0)} \tag{23}$$

Based on the calculation result of the signal-to-noise ratio, several low-frequency noise components are removed to ensure that the remaining low-frequency noise performance is significant [23], which is equivalent to filtering by a high-pass filter. The filter bank is completely adaptive, and the high-pass filter can be expressed as:

$$x_h(t) = \sum_{j=1}^k c_j(t) \tag{14}$$

In the formula,  $j = 1, 2, \dots$ , which means to retain the first  $k$  IMF components of the signal, so that a high-pass filter can be constructed. Retain the significant performance of low-frequency noise, and transfer high-frequency energy to low-frequency.

Assuming the center frequency  $f(m)$  of filter, the frequency response of the filter is  $H_m(k)$ :

$$H_m(k) = \begin{cases} 0, k \leq f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, f(m) \leq k \leq f(m+1) \\ 0, k > f(m+1) \end{cases} \tag{25} \quad 0 \leq m \leq M-1$$

In the formula, the center frequency  $f(m)$  of each filter is:

$$f(m) = \left(\frac{N}{f_h}\right) B^{-1} \left( B(f_l) + m \left( \frac{B(f_h) - B(f_l)}{M + 1} \right) \right) \tag{26}$$

In the formula,  $f_h$  and  $f_l$  are the highest and lowest frequencies of the filter bank,  $f_s$  is the sampling frequency,  $N$  is the number of sampling points per frame, and  $M$  is the number of filters.  $B^{-1}$  is the inverse function of  $B$ .

The energy spectrum  $P(f)$  is passed through the Mel filter bank. The passable frequency range of this wave device group is  $0 \sim f_N/2$  ( $f_N^-$  is the sampling frequency).

$$E(m) = \sum_{k=0}^{N-1} p(f) \cdot H_m(f) \tag{27}$$

Among them,  $N$  is the total number of points of each frame signal, and  $M$  is the number of filters.

From the design idea of the Mel filter, it can be known that when the center frequency of a certain Mel filter is  $f(m)$ , the frequency response function of the filter is  $H_m(k)$ .

$$H_m(k) = \begin{cases} 0 & k \leq f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \tag{28}$$

In the formula,  $0 \leq m \leq M-1$ , the number of filters is  $M$ , and the calculation formula for the center frequency  $f(m)$  of the Mel filter is:

$$f(m) = \left(\frac{N}{f_s}\right) Mel^{-1} \left( Mel(f_l) + m \frac{Mel(f_h) - Mel(f_l)}{M + 1} \right) \tag{29}$$

In the formula,  $f_s$  is the sampling rate,  $f_l$  and  $f_h$  are the lowest and highest frequencies of the filter bank respectively,  $N$  is the number of sampling points of each frame of data, and  $Mel^{-1}$  is the inverse function of  $Mel$ .

Assuming that  $x(t)$  is the most stable low-frequency noise performance of electronic components, it is assumed that the correlation function  $R_x(t, \tau)$  of  $x(t)$  is a periodic function with  $T$  as the period in the time delay domain. That is  $R_x(t, \tau) = R_x(t + T)\tau$ . Then  $F$  is expanded in the form of  $R_x(t, \tau)$  ourier series, and the low-frequency noise performance extraction results of electronic components are:

$$R_x(t, \tau) = \sum_{\omega=-\infty}^{\omega=\infty} R_{x\omega}(\tau) \cdot e^{j2\pi\omega t} \tag{30}$$

## 4 Analysis of Results

To verify the application performance of the low-frequency noise extraction method of electronic components.

### 4.1 Basic Settings

The low-frequency noise performance of electronic components is selected for reconstruction, and the result is shown in Fig. 2 below.

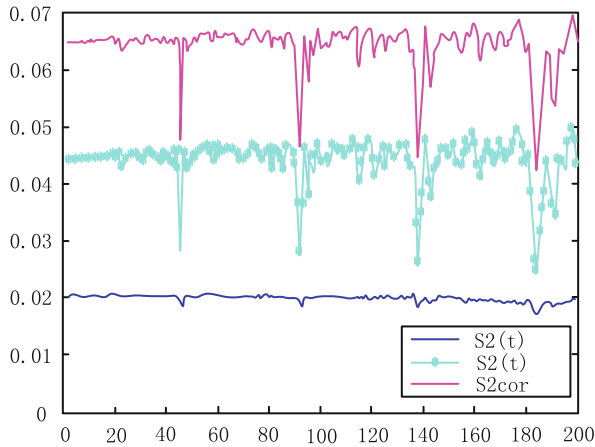


Fig. 2. Low-frequency noise performance reconstruction

In the above figure, the horizontal axis  $t$  represents the time variable, and the three different colored curves from top to bottom represent the values of the relevant index  $S_{2cor}$  of the statistic, the difference  $\Delta\bar{S}_2(t)$  of the statistic, and the statistic  $\bar{S}_2(t)$ . The time delay parameter is determined by the first zero point of  $\bar{S}_2(t)$ , or the time corresponding to the first minimum value of  $\Delta\bar{S}_2(t)$ , and the time window  $\tau_w$  is determined by the time corresponding to the minimum value of  $S_{2cor}$ . Phase space is the model basis for the extraction of chaotic characteristics and the basis for the extraction of low-frequency noise performance of electronic components.

### 4.2 Comparison of Extraction Accuracy

In order to make the proposed method effective, the following experimental tests further compare the accuracy of different extraction methods. The specific experimental comparison results are shown in Table 1:

**Table 1.** Accuracy comparison of different extraction methods

Number of test samples/ (a)	Accuracy/(%)		
	The proposed method	Literature [5] method	Literature [6] method
15	0.958	0.508	0.652
20	0.924	0.578	0.674
25	0.945	0.557	0.685
30	0.974	0.595	0.602
35	0.986	0.512	0.754
40	0.935	0.604	0.771
45	0.957	0.634	0.774
50	0.964	0.645	0.781

The analysis of the experimental data on Table 1 shows that the accuracy of the literature [5] method reaches 0.645, and the literature [6] method is 0.781, while the accuracy of this method is up to 0.964,. It is seen that this method effectively solves the problem of low-frequency noise performance extraction of electronic components, effectively avoids external interference, effectively improves the accuracy, and significantly outperforms the other two extraction methods. The superiority of the extraction results of the proposed method is fully demonstrated.

## 5 Conclusion

To solve the problem of low extraction accuracy in traditional extraction methods, a data mining-based low-frequency noise performance extraction method of electronic components is proposed. By introducing data mining methods, the analysis of low-frequency noise performance can be effectively improved, and the improved method is effectively improved by combining modulation and filtering. The effectiveness of extraction and the significant increase in extraction accuracy have certain advantages.

## References

1. Wang, F., Ni, M., Zhou, M., et al.: The optimal solution of quantum adiabatic approximation for maximum cut problem. *Comput. Eng.* **509**(01), 31–36 (2020)
2. Wu, S., Luo, Y., Chen, J.: Steady state characteristics of power function single well system driven by multiplicative and additive noise of color correlation. *Vibration Shock* **39**(5), 244–249 (2020)
3. Chen, J., Liang, Z., Fu, C.: Research on speech noise power spectrum estimation algorithm based on time recursive average. *J. Sichuan Ordnance Eng.* **040**(001), 135–139 (2019)
4. Zhu, M., Duan, Z., Guo, B.: Noise reduction effect analysis of vibration screen bearing signal based on EEMD and wavelet packet. *Mech. Design Manuf.* **351**(05), 70–74 (2020)

5. Geng, Y., Zhang, R.: Channel equalization algorithm for MIMO filter banks based on SINR. *Sci. Technol. Eng.* **478**(09), 138–142 (2019)
6. Xia, W., Cai, Z.: Performance analysis of power spectrum entropy detection for unknown sinusoidal signal. *Signal Process.* **34**(12), 109–117 (2018)
7. Wang, M., Liu, Y.: Partial discharge ultrasonic signal monitoring and digital noise reduction. *Mech. Manuf. Autom.* **258**(05), 208–211 (2018)
8. Du, F., Tang, L.: Convergence analysis of frequency spectrum and Fourier series of unit impulse sampling sequence. *Vibration Shock* **38**(04), 20–24 (2019)
9. Liu, H.: Order analysis method of nonstationary exhaust noise signal based on wavelet transform. *Vibration Shock* **038**(022), 29–35, 51 (2019)
10. Chen, P., Xiang, S., Li, M., et al.: Frequency estimation of multi-stage signals based on linear predictive autocorrelation. *Sensors Microsyst.* **336**(02), 63–65 + 69 (2020)
11. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>
12. Liu, S., Liu, X., Wang, S., Muhammad, K.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
13. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
14. Liu, H., Ran, H.: Simulation of medical data information extraction based on data mining technology. *Comput. Simul.* **37**(05), 375–378472 (2020)
15. Peng, Y., Tao, S., Peng, T., et al.: Extraction method of parallel arrangement entropy characteristics of time series data under cloud platform. *Power Autom. Equipment* **39**(04), 217–223 (2019)
16. Wang, D., Xu, T.: Low frequency generation and compound noise behavior in blue violet light emitting diodes. *Acta Physica Sinica* **68**(012), 273–278 (2019)
17. He, Y., Liu, Y., Zhang, X.: Analysis of hot carrier injection effect in 65nm process NMOS devices based on low frequency noise. *Semiconductor Technol.* **44**(07), 50–55 (2019)
18. Dong, S., Guo, H., Ma, W., et al.: Ionizing radiation damage mechanism and bias dependence of AlGaIn/GaN high electron mobility transistor devices. *Acta Physica Sinica* **69**(07), 294–302 (2020)
19. Luo, Z., Yan, H.: Robust laser micro Doppler feature extraction method based on pca-clean. *Laser Infrared* **50**(11), 27–35 (2020)
20. Hu, Q.G., Zhou, L.: Analysis of secondary transmission path of low frequency noise in car compartment and noise reduction control. *Noise Vibration Control* **40**(01), 80–85 (2020)
21. Sun, Z., Ye, Q.: Physical unclonable function for extracting reliable information entropy of multiple bits by using frequency characteristics of oscillating ring. *Acta Electronica Sinica* **43**(01), 234–241 (2021)
22. Jia, X., He, L.: Low frequency noise test method for microwave integrated circuits. *Semiconductor Technol.* **44**(02), 67–71 (2019)
23. Chen, L.: Anechoic properties of acoustic metamaterials for low frequency noise. *Appl. Acoust.* **39**(03), 438–444 (2020)



# X-band Radar Detection Target Tracking Method Based on Internet of Things Sensing Technology

Fengshuo Yan<sup>(✉)</sup>, Kui Xiong, Mingyang Gao, and Yanxin Shu

The Second Research Institute of CAAC, Chengdu 610041, Sichuan, China  
Yfs215510@163.com

**Abstract.** If echo data of civil aviation wake scattering can be obtained by Aircraft wake X-band radar detection, which can provide reference data for the stable operation of aircraft. However, the existing target tracking methods have the problem of imperfect target tracking mode, which leads to poor echo strength. A target tracking method of Aircraft wake X-band radar detection based on IOT sensing technology is designed. The structural characteristics of dynamic target are acquired, the target tracking task is formulated, the Aircraft wake X-band radar detection imaging area is set, the distance between the detection imaging area and the target position is calculated, the information fusion technology is used to suppress the same frequency interference clutter, the radio frequency rate of radar in frequency domain and space domain is adjusted, and the sliding window residual is used to obtain the track extrapolation motion track, and the target distance and the target position are measured. According to the spatial position of azimuth angle and pitch angle, the target tracking mode is set based on the Internet of things perception technology, and the data association rules are designed. Experimental results: the average echo intensity of the designed tracking method is 16.777% and 15.461% higher than that of the two existing target tracking methods, respectively, which proves that the target tracking method combined with IOT sensing technology has better practical application effect.

**Keywords:** Internet of things perception technology · Aircraft wake X-band radar · Tracking method · Echo intensity

## 1 Introduction

Aircraft wake is the inevitable product of aircraft lift, its vortex strength is large, lasting for a long time, which has a great impact on the aircraft into which, in serious cases, can lead to aircraft crash. In order to avoid the occurrence of such wake accidents, acoustic, electrical, optical and other detection equipment can be used to detect the aircraft wake, and provide the position, strength and other information of the wake, so as to further enhance flight safety, break through the limitation of safe flight distance standard formulated by the international Civil Aviation Organization (ICAO), and improve transportation efficiency. Due to the advantages of radar in frequency selection



and detection range, more and more attention has been paid to the radar detection technology of aircraft wake.

The traditional detection before tracking technology uses the process of detection before tracking to detect and track targets. After beamforming and pulse compression preprocessing, single frame radar target echo data initially accumulates target energy. Through coherent or non coherent signal processing, the target echo energy is further accumulated, and the detection threshold is set according to the false alarm rate required by the system to detect the echo data, then the parameters of the detected point trace are estimated, and finally the processing results are transmitted to the tracking system to complete the target tracking [1]. In the traditional signal processing flow, the radar echo data is processed by CFAR detection. This operation can reduce the amount of data, but from the perspective of information theory, the amount of information of the target in the data is lost, resulting in the loss of signal-to-noise ratio of the target echo. When CFAR is used to detect weak targets, there is a trade-off between the probability of target detection and the probability of false alarm. In order to ensure a certain detection probability, a lower detection threshold is usually selected. Therefore, in single frame detection, the target may have multiple detection points, resulting in the increase of false alarm number and the uncertainty of observation source. The non target observation with wrong tracking selection can lead to the divergence of tracking filter. Therefore, the correct association of the target observation click with the target track has an important impact on the target tracking accuracy. Target tracking refers to the use of sensors to observe and determine the location, trajectory and characteristics of the target. The target tracking problem faced by modern military is extremely challenging. In the complex environment of modern battlefield, the characteristics of the target such as strong mobility and low RCS make the traditional target tracking method not only affected by the unknown number of targets, but also the low scattering cross section (RCS) The impact of clutter density and other problems, but also increased the low detection probability, data association difficulties and other new problems of interference, the target tracking algorithm put forward higher requirements. Modern radar system requires a tracking algorithm to detect the target track and track it accurately in a complex environment, so as to provide accurate information for the command system and fire control system, so as to effectively interfere with the interested target and reduce the false alarm probability as much as possible while ensuring high detection probability. The final effect of tracking algorithm is the embodiment of radar comprehensive performance. On the one hand, it can improve the performance of tracking algorithm. On the other hand, it should also improve the quality of point trace input as much as possible. The echo intensity of traditional early warning radar to the target has the characteristic of "large near and small far"; The interference of observation noise and clutter is added; In addition, the RCS of high frequency radar cross section is sensitive to the change of target attitude. The radar detection probability is low, and sometimes it can't detect the target effectively for several consecutive cycles [2]. Especially with the development of modern weapon technology, a series of measures, such as electronic jamming, aircraft stealth technology, anti radiation missiles and so on, pose challenges to radar warning, so the combat intelligence support system based on single radar can not meet the needs of the battlefield.

With the improvement of economic level, the upgrading of hardware platform and the development of information fusion technology, multi radar cooperative detection technology provides an effective way to deal with the increasingly complex war environment. Radar detection is not only a new early warning detection theory and method, but also provides advanced and effective equipment and tactical support for radar reconnaissance troops to carry out cluster detection and system operations. Compared with single radar, multi radar has the following advantages:

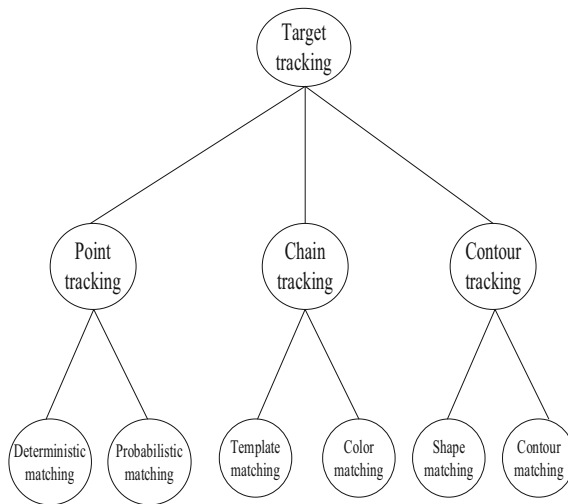
1. Expansion of detection range and target types; The detection range of multiple radars expands with the expansion of the distribution range of each radar. The spatial diversity greatly improves the scintillation of the target echo. The characteristics of multi angle observation target, especially the tracking interruption caused by single radar over the top and the detection blind area caused by electromagnetic environment, are compensated. Because the bands of multiple radars may be different, frequency diversity improves the types of detected targets and greatly improves the defense capability of anti radar technology against enemy electronic jamming.
2. Improve system reliability: multi radar system can maintain each radar of multi radar one by one in non wartime without reducing detection coverage. In wartime, even if a radar breaks down suddenly, multi radar system will retain certain warning ability.
3. Improvement of target tracking accuracy and recognition ability: different radars of multi radar system may achieve multiple coverage of a certain target in a certain area. Multi band coverage of the same target can not only verify the data of multiple radars, but also improve the confidence of the target, Through the configuration between radars, the target detection accuracy of multi radar system can be improved, and even different information can be obtained from each radar to improve the target recognition ability.

## **2 Aircraft Wake X-band Radar Detection Target Tracking Method Based on Internet of Things Sensing Technology**

### **2.1 Acquisition of Dynamic Target Structural Features**

Dynamic target tracking can be defined as estimating the target's trajectory in a scene. However, in the process of tracking, the target may change in occlusion, deformation, scale and illumination. This undoubtedly brings great challenges to the tracking task. To effectively solve the difficulties in the task, the key lies in the following points: first of all, it is to find a reasonable method to mark the target and describe the target with appropriate features [3]. Secondly, find an effective method to track and detect the target, and update the characteristics of the changed target in time. In the tracking method, points can be used to represent the object. Point is used to represent the target, which requires the target to have the key features that can be represented by point, such as the center of mass of the object, the joint point of motion. The targets that can be represented by points generally occupy a small area in the image, such as the moving

targets such as shells and birds in the air, vehicles and pedestrians in the distance. A rectangle and a circle frame are used to delineate the target, which represents the target to be tracked. This method is a common form of target representation in current tracking methods, and it has the characteristics of image, simplicity and so on by using simple geometry [4, 5]. It is mainly suitable for rigid targets, but it is also used for tracking flexible targets, such as pedestrians, animals and so on. However, it puts forward strict requirements for feature selection. In the process of target tracking, the most intuitive image perception of the observer is the contour of the target. Through the contour to find the edge of the target and the background, the area within the contour is the tracked target. The use of contour to mark the target has the characteristics of accurate and intuitive. However, when the edge of the target is not clear enough, the contour information cannot be obtained. In the video tracking task, the application of human tracking is very extensive, because the human body can be regarded as a target connected by joints, which forms a chain region. With this chain connected region, the human body is calculated by calculating the motion model between the parts, and represented by multiple ellipses. For the rigid target or the target with the characteristics of chain region, the target can be represented by skeleton structure. People or animals can usually be represented by the same backbone structure. For tracking tasks, the backbone structure focuses more on reflecting the motion information of each part of the target. For the diversity of tracking tasks, multiple representation methods can also be used to represent the target together. The specific structure is shown in Fig. 1



**Fig. 1.** Target tracking structure

As can be seen from Fig. 1, structural features are obtained through deep neural network training, which are structural features from simple to complex. Deep neural network refers to the structure of animal visual cortex. Taking humans as an example, the information processing of human visual system is hierarchical. From the lower level

to the middle level, the shape features are formed by the combination of edge and corner features, and then to the higher level, even to the whole target. It can be said that the process of visual perception is a process of abstracting from the underlying basic features to the high-level structural features. As the level of abstraction becomes higher and higher, the displayed features have higher-level representation. In the face of the changing targets in the tracking task, the features from the bottom to the high-level can be used to describe [6]. Color features are widely used. In image processing, with the intuitive RGB color space, easy to calculate, widely used to mark the target color. Color histogram is a common color feature, which calculates the proportion of each color in the image color space instead of calculating the position of each color in the image space. For the target without clear edge in space, the description is strong. But for the judgment of color similarity, human subjectivity plays a decisive role, so some people put forward HSV space based on image chroma, saturation and brightness value, which makes it closer to human visual subjective recognition of color through the description of color features [7, 8]. The conclusion that the visual cognitive process of human vision to the target is produced by the movement of the edge can be seen from the experiments of the potential visual mechanism. When the target edge changes, the edge between the object and the background will change dramatically. In the process of setting tracking task, edge detection method can be used to detect the change of edge. Compared with color features, the generation of edge features depends on the moving edge generated by the target and background, not on the subjective understanding of color. So its simplicity and accuracy can be guaranteed.

## 2.2 Set up the Aircraft Wake X-band Radar Detection and Imaging Area

Setting the Aircraft wake X-band radar detection imaging area is to match the corresponding points in the front and back video frames, so as to complete the target trajectory tracking. Due to the use of point based features, when the target's rotation, scale and affine change, the point matching will not be affected. However, once the target is occluded and the illumination changes, the target will disappear or appear, will directly affect the imaging area. In continuous video frames, feature points are matched before and after frames, and multiple matching relationships are solved by minimizing the matching cost according to the motion constraints of feature points. So as to find the best match. The principle of optical flow method is to calculate the motion vector of the relative moving object by matching the corresponding points in the front and back frames to form the optical flow field. The Aircraft wake X-band radar can reach  $360^\circ$ . At present, the targets of continuous circular scanning in all directions have been widely installed on ships to avoid collision and ensure the safety of people and goods. The working principle is that the transmitter transmits a series of high-frequency pulses which are repeated in a certain period to the observed direction through the antenna. If there is a target in the propagation path, the target will intercept part of the electromagnetic energy emitted by the radar and radiate the intercepted energy again in the form of electromagnetic wave. Part of the scattered energy will face the direction received by the radar (called target echo). The radar antenna can receive this part of the energy scattered by the target and process it, display the image on the screen. After acquiring radar image data, preprocessing is needed. It includes the determination of

the analysis area for subsequent inversion, data correction, image filtering and coordinate transformation. In signal analysis, if the amount of data used for analysis is too small and the amount of information contained in the signal is insufficient, the real information cannot be restored. However, if too much data is used, the statistical characteristics of the data will be different, which will affect the accuracy of the results. Tracking target can be seen as the superposition of countless sine waves. The acquisition of target information is actually the acquisition of countless sine signals. This paper reasonably selects the amount of data required for each analysis, so that the radar image analysis also has fixed characteristics. In addition, the goal is usually considered as a stationary ergodic stochastic process. By selecting enough data from the samples, the time domain and frequency domain characteristics of the whole system can be analyzed. Therefore, in the actual inversion, several consecutive clutter images are used each time, and the time length is about 1 min. According to the intensity distribution of radar echo image, the distance between the detection imaging area and the target position is calculated. The specific expression formula is as follows (1)

$$Q = \frac{E}{\beta \tan\left(\frac{\delta\pi}{9}\right)} \quad (1)$$

In formula (1),  $E$  is the radar erection height. When the location of the selected area and the radar detection imaging area is 750 m, the number of pixels is 85, and the value range of the center angle is  $60\text{--}70^\circ$ . The polar coordinate expression formula of the imaging region is as follows:

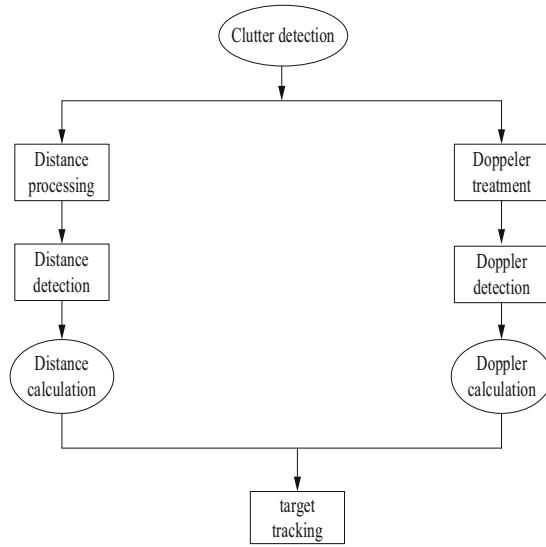
$$\begin{cases} a = l \times \cos\delta \\ b = l \times \sin\delta \end{cases} \quad (2)$$

In formula (2),  $a, b$  represents any two points in the radar imaging area, and  $l$  represents the entire point function value. Radar sea clutter image contains a lot of noise jamming, such as target jamming, rain and snow jamming, CO frequency jamming and so on. The signal needed is weak. Therefore, it is necessary to use median filter to process the image of the region simply and filter out the interference of white noise [9]. When detecting and tracking the imaging area of the target in motion, the constant change of motion attitude and high-speed motion will increase a Doppler frequency to the radar echo. Therefore, it is necessary to compensate for velocity and attitude. Firstly, 32 consecutive images are compensated by angle motion. The first image is taken as the reference, and the other 31 images are rotated accordingly to ensure the same angle to the ground as the first image. Secondly, position correction is performed on 32 consecutive images in the same way [10, 11]. Frequency aliasing is a phenomenon in the process of signal acquisition. With the advent of the digital signal era, the use of computer processing signal data has been recognized by the majority of technical workers. However, because the computer can only process digital signals, it is necessary to sample analog signals and convert them into digital signals. Sampling theorem is the basis of understanding frequency aliasing. The basic meaning of the

sampling theorem is: when the signal is sampled discretely, the sampling frequency should be greater than twice of the highest frequency of the signal, so that the sampled digital signal can completely retain the information in the original signal. The radar antenna transmits electromagnetic wave and receives its echo every 100 rotations. The energy distribution of each line is recorded in the image data. The radar scans its area in 3600 directions, so that the radar image contains a lot of information, which does not meet the needs of the algorithm. Therefore, the data in the above rectangular box is still used for wave inversion. After determining the center angle and starting position of the selected rectangle, the nearest point interpolation is used to obtain the imaging area of radar detection.

### **2.3 Inhibition of Clutter by Information Fusion Technology**

Target location is one of the main functions of radar. As an active detection sensor, the radar transmits electromagnetic waves to the monitored field through the transmitting antenna, and then receives all backscattered echoes from the objects in the field through its receiving antenna. By processing the echo, the feature information of the interested target is extracted, and then the state information of the target, such as presence, position, motion state, friend or foe attribute, is obtained. The position information of the target can be obtained from the observation results of the arrival time, the angle of arrival, the time difference of arrival and the received signal strength of the echo [12–14]. For the traditional location method based on direct wave information, each radar station can only obtain a single characteristic information of the target location, whether it is a monostatic radar with the same receiving and transmitting antennas or a bistatic radar with separate receiving and transmitting antennas. Because the distance between the same frequency jamming radar and the experimental radar is too close, and the jamming clutter is a very strong signal, the echo formed by the same frequency jamming noise in the radar image is very saturated. In general, it is more difficult to eliminate the same frequency synchronous jamming than the asynchronous jamming. In general, the elimination method is to use some suppression methods to suppress the same frequency synchronous interference after it is converted into asynchronous interference. In the experiment, the same frequency interference in the radar image collected is basically asynchronous interference, so the subsequent detection and suppression methods are all asynchronous interference. In order to suppress the co channel interference, the active interference suppression method can be used. First of all, the signals transmitted by radars can be distinguished or unique by certain methods, which is also the most fundamental method to suppress co channel interference. Firstly, the same frequency interference clutter is detected, and the distribution range is determined before centralized processing. The detection process is shown in Fig. 2



**Fig. 2.** Same-frequency interference detection process

As can be seen from Fig. 2, in the process of target tracking, it is necessary to reduce or even avoid the same frequency interference energy received by the radar receiver. It can adjust the transmitting frequency of radar in frequency domain, spatial domain or time domain, interleave the working time of the same type of radar, or extend the distance between the formed ships, so as to minimize or avoid the same frequency interference received by radar receiver. The echo characteristics of the monitored target are different from those of the same frequency interference to achieve the purpose of suppressing the radar receiver receiving the same frequency interference [15, 16]. Firstly, the number of repeat frequency points can be increased, then the same frequency synchronous jamming can be converted into asynchronous jamming by certain methods, and finally the corresponding signal processing methods can be used to reduce the influence of the same frequency jamming. Adjust the radar receiving frequency. The common processing methods are: median filter, mean filter, adaptive filter, geometric filter based on morphology, filter based on wavelet transform and so on. These methods have good results, but the first several classical methods will change the target information in the radar image; The latter two methods are slightly better in preserving image information. The most widely used wavelet transform is to remove the Gaussian noise in the image, and its effect is very good. For bistatic radar, the distance traveled by the signal between the radar and the target is not equal. Therefore, if the bistatic radar obtains a certain arrival time, the point trace of the possible position of the target meeting the arrival time forms an elliptical contour with the transmitter and receiver as the focus, and the sum of the distance between the target and the transmitter and the receiver as the focus [17]. According to the difference of pulse repetition period and transmitting frequency, the same frequency jamming of radar can be divided into

synchronous jamming and asynchronous jamming. If the repetition frequency and pulse width of the pulse transmitted by one of the radars are known, and the repetition frequency and pulse width of the pulse transmitted by the other radar are also known, the relationship between them is expressed as follows:

$$\Delta d = \frac{1}{k_1} - \frac{1}{k_2} \quad (3)$$

In formula (3),  $k_1, k_2$  represents the pulse frequency transmitted by two radars. At the same time, the more the radars are turned on, the more the concentric circles are. The width of the same frequency synchronous jamming is generally larger than the corresponding distance of the radar pulse width, and the concentric circles will move outward (inward) slowly. In addition, due to the slight difference in the period of each radar, the radius of the jamming loop on the radar screen changes with time: the display of one radar changes from small to large, and changes circularly; The other part may change from large to small. If more than two radars work at the same time, the pattern of image jamming will be more complex. The image of the same frequency asynchronous jamming on the radar screen is fan shaped and divergent dotted line, and the concrete manifestation of the same frequency asynchronous jamming on the radar screen is not unique [18, 19]. With the difference of interference pulse width and working repetition frequency, the dots of the scattering fan shape are different in length. Usually, when the display works in a large range, the interference source is close to each other, and the antenna direction is right, the interference becomes serious. Due to the random range and azimuth of the interference, especially when the interference density is high, it has a certain impact on the tracking stability.

#### 2.4 Obtain the Orbital Extrapolated Motion Trajectory with the Sliding Window Residue

In the process of target tracking, the design of tracking filter is a very important problem. Using sliding window residuals to obtain trajectory extrapolation requires tracking filter, which can obtain more accurate measurement results by jointly processing the current and past measurement data. At the same time, the filter predicts the position of the target at the next moment, and the radar controls the tracking gate and antenna (beam) pointing according to the prediction results. It can be seen that tracking filtering is the basis of continuous measurement [20]. Least square filtering is a very robust method because it does not make any assumptions about the distribution of data. This characteristic makes it widely used in tracking filtering. The improved form of least square filtering: sliding window residual least square filtering. The main idea is to eliminate the main nonlinear term in the observation data by orbit extrapolation, and use a simple polynomial model to smooth the residual. Through the sliding window method, high-precision final results can be obtained without changing the data rate. In this process, it is assumed that the earth is a uniform sphere, but in fact, due to the influence of rotation, the earth is an ellipsoid, and its equatorial radius is 21 km more than the polar radius. As a result, the gravitational field of the earth does not point to the center of the earth. The calculation formula of orbit extrapolation is as follows:



$$w(f + v) = z(v) + f \times \frac{s_1 + s_2 + s_3 + s_4}{4} \quad (4)$$

In formula (4),  $w$  represents the time of the orbit extrapolation,  $f$  represents the semi-long axis of the track,  $v$  is the eccentricity of the track,  $z$  is the inclination of the track, and  $s$  is the included angle of the earth center. Similar to the state vector of the target, the orbital root number also describes the motion of the space object completely. The position and velocity of the target at any time can be obtained by knowing the number of orbital roots of the target; On the contrary, the number of orbital roots can be obtained by knowing the position and speed of the target at a certain time. At present, some foreign scholars publish the root number information of all the catalogued targets in the form of TLE file every day [21, 22]. The title data consider the factors of various perturbations including the earth's non spherical, solar and lunar gravity, atmospheric resistance, and eliminate the periodic influence by specific means, so it is an average root number. When using the title information to extrapolate the orbit, the same method must be used to reconstruct the periodic disturbance term to obtain the accurate position and speed of the target. In engineering application, TLE data is usually used in combination with a gp4 model to extrapolate the track to achieve the best prediction accuracy. Radar can obtain its position in space by measuring the distance, azimuth and pitch angle of the target. Radar measurement always has errors, including physical error and measurement error. Generally, the physical error can be compensated by calibration in advance, so that the main source of measurement error is random noise. Filtering is a method to reduce the influence of random noise [23]. By filtering the measurement with random noise superimposed, the relative more accurate measurement can be obtained. The simplest linear model is polynomial model, but the radar observation equation is highly nonlinear. The mismatch between the filter model and the actual situation seriously affects the accuracy of the filter, and even introduces additional physical errors. Although the higher order polynomial can be used to approximate the nonlinear real situation, higher-order model means more computation. The results of the second order fitting and the third order fitting are not different from those of the third order fitting. The maximum error (absolute value) of fitting is about 2 m, considering the distance of the target is about 600 km, so the fitting error of this degree can be ignored. For azimuth, the second order fitting effect is very poor, and its maximum error (absolute value) reaches about  $10^\circ$ , while the third order fitting has obvious improvement, but its maximum error (absolute value) is about  $2^\circ$ . Similar to the distance, the fitting results of pitch angle are also ideal. The difference between the second and third order is not big, and the maximum error (absolute value) is about  $0.2^\circ$ . It is predicted that the maximum azimuth error of the third order polynomial least square filter is about  $2^\circ$ , which will directly lead to the loss of the target. Since the reference orbit is not far from the real orbit of the target, it can be considered that the nonlinear term contained in the reference orbit is consistent with the non-linear in the observation data. By subtracting the measured data and reference orbit, the main nonlinear terms in the observed data are eliminated. The residual data will mainly include the residual error of the orbit and the random error of the measurement. Compared with the original observation data set, the residual error has been greatly reduced, so it is possible to use low order polynomial fitting to reduce the influence of

random error. Some catalogued targets can get the information of orbital elements in advance, so it is inferred that the extrapolated orbit can be used as reference orbit. However, space target is a non cooperative target, especially for ballistic missile, it can not obtain the previous information in advance. So the more feasible method is to estimate the state of the target in real time in the process of filtering, and the orbit derived from the extrapolation is used as the reference orbit.

## 2.5 Setting Target Tracking Mode Based on Internet of Things Perception Technology

Based on the perception technology of Internet of things, target tracking mode is set. The perception layer is a more complex part of the trust mechanism in the Internet of things, mainly because of the heterogeneity and openness of these two factors: the first factor is due to the lack of a unified standard, now the implementation of information collection and transmission is mainly implemented by the industry using different hardware, resulting in many different application models. At the same time, different terminals have different processing capabilities, which makes them diversified, so they can not rely on a single trust mechanism. Sensor nodes can be deployed either through the interaction of sensor nodes or through the access of sensor nodes. At the same time, the interaction of data is also very strong, and even can interact across institutions. The fuzzy boundaries are divided into clear boundaries, so that the perception data can find its trust level, find its corresponding reputation organization, and authorize it to obtain the corresponding trust value, so as to make the complex perception data orderly. In fact, the perception layer is not a physical layer. By definition, the perception layer is defined as a hierarchical mechanism, so that each layer has its own different trust needs. When the perception data flows gradually in each layer and meets the trust needs of each layer, the perception data is safe and effective, This also ensures the availability of sensing data, rather than a pile of redundant and useless garbage data. Things are connected, it is not limited to the computer network, but a number of heterogeneous networks connected with each other, its data has been the previous power exponent times There is a problem: how to separate their own useful data from the mass data, how to determine the legitimacy of these data, people need to design the corresponding association rules. In order to avoid the congestion of Internet of things nodes, the fairness of network nodes is measured.

$$h = \frac{\left(\frac{r_1}{p_i}\right)^2}{D\left(\frac{r_2}{p_i}\right)^2} \quad (5)$$

In formula (5),  $r$  is the total number of network nodes,  $D$  is the number of packets,  $p$  is the average packet rate, and  $i$  is the energy consumption coefficient of data transmission. Considering the attenuation of data, the perceptual data is endowed with certain vitality. Due to the real-time nature of the data on the Internet of things, not all

data is immutable. On the contrary, data can be updated immediately. After a period of time, in the time definition of the Internet, the time taken to change is very short. These perceptual data will take on other forms after changes. Through the relevant algorithm to determine whether it is near the limit of life, if there is, then discard the data, think it is useless. It mainly determines whether the reputation value of perceptual data is rapidly declining, or malicious data. If the perceptual data is considered useless or even threatening, it will be rejected and discarded. In the implementation of specific operation protocol, it is above the multiplication gate and the addition gate. Analog operation, pseudo-random operation and polynomial operation are common problems. From a more high-end point of view, it belongs to the field of computing related to geometric computing, statistical analysis, data mining, special operations, such as linear algebra operations, privacy preserving data query and pattern recognition, etc.; From a more high-end point of view, it can be some specific agreements, such as electronic like, electronic purchase auction and so on. In the complex environment, the echo signal returned from the target reaches the radar antenna through different paths, which reduces the accuracy of target positioning and even causes the problem of false target. Generally speaking, for the point target with large elevation angle, the multipath component enters from the radar antenna sidelobe, and the monopulse radar can locate the target with high accuracy. However, for low elevation point target, due to the multipath signal entering the sum and difference main beam, there will be large positioning error, and even lead to positioning failure. At this time, if the traditional matched filtering method is used to locate, due to the interference of multipath signal, the radar will mistakenly detect multiple false targets, so it is unable to accurately detect the real position of the target. Under the condition that the geometric characteristic parameters of buildings and roads in the detection area are clear, the multipath propagation can be predicted. After geometric transformation, the detection results of mirror position points can be converted into the results of real observation position, so the multipath echo also has available information. The calculation formula of echo signal received by receiving antenna is as follows:

$$g = \sum_{b=1}^3 \Gamma_b^2 c (m - \tau_b) \quad (6)$$

In formula (6),  $b$  is the arrival time of the echo,  $\Gamma_b$  is the echo amplitude,  $c$  is the data noise,  $m$  is the echo path, and  $\tau_b$  is the echo velocity. From the perspective of inverse ray tracing, it can also be considered that multipath propagation introduces the virtual mirror observation base station. If the multipath information can be used effectively, then the target location can be achieved only by a single radar station. Combined with the technical characteristics of IOT sensing technology in practical application, the design of Aircraft wake X-band radar detection target tracking method is realized.

### 3 Experimental Study

Through prior knowledge and association rules, target tracking in low SNR scenes can be achieved. However, this algorithm considers all candidate echoes in the correlation gate comprehensively, and the tracking error is larger in clutter dense scenes than in ideal detection conditions using only the correct observation of Kalman filter. In order to verify the radar echo strength of the designed target tracking algorithm in different SNR scenes, two existing tracking methods are selected and compared with the designed tracking method. In this paper, the effectiveness of the proposed method is verified by the X-band radar detection experimental data of a certain type of civil aircraft wake.

#### 3.1 Setting Experimental Parameters

According to the needs of experimental test, the experimental parameters are set as shown in Table 1.

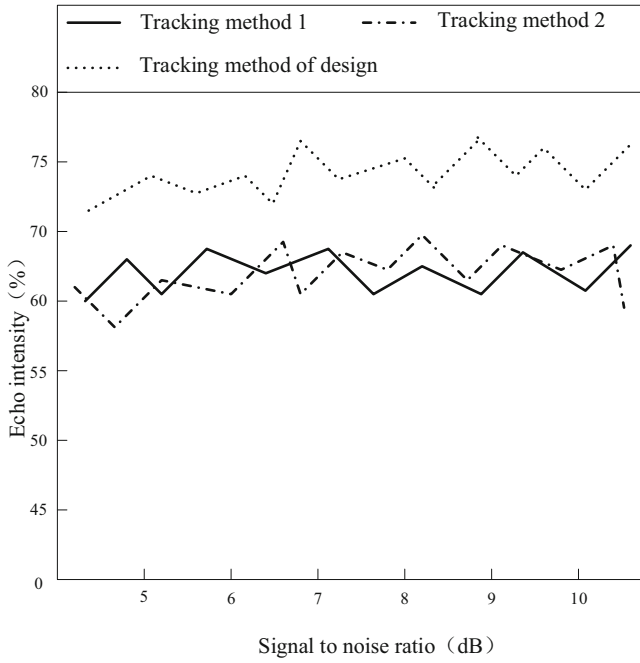
**Table 1.** Tracking method smoothing parameters

Serial number	Name	Parameter
001	Carrier frequency	3.5 GHz
002	Bandwidth	4 MHz
003	Pulse repetition rate	1.92 kHz
004	Number of coherent integration pulses	34
005	Half power beamwidth	8°
006	Observation interval	12 s
007	Detection probability	0.88
008	Clutter density	$10^{-6}$
009	Correlation gate threshold	18.2
010	Gate probability	0.9964
011	Fixed interval length	10 frames

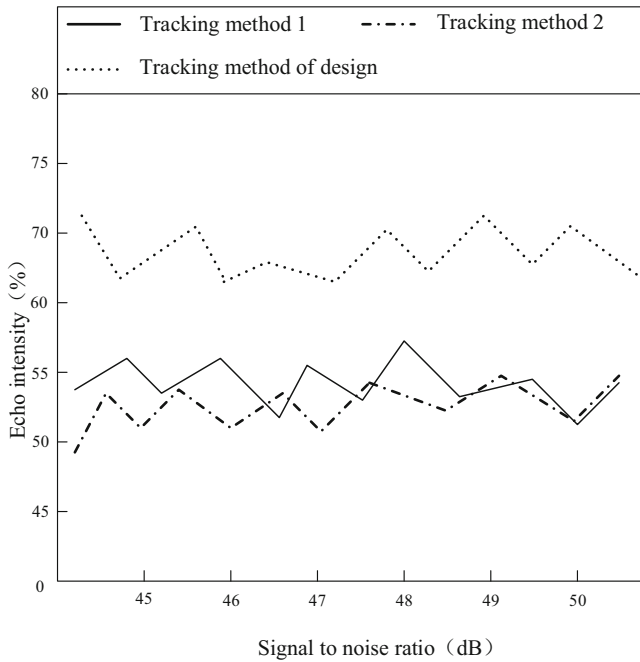
Under the above experimental parameters, the echo intensity of the three target tracking methods is tested, and the experimental results are obtained.

#### 3.2 Experimental Results

Suppose that the X-band radar peak power of the civil aircraft wake is  $10^6$  W, the transmission gain is 40 dB, the wavelength is 0.056 m, the pulse accumulation number is 16, the radar cross section is  $3 \text{ m}^2$ , and the receiver's direct frequency bandwidth is  $1.6 \times 10^6$  Hz. The minimum output signal-to-noise ratio of the receiver is 2, and the noise temperature of the radar receiver is 290 K. Under different SNR conditions, the echo strength of three tracking methods is tested, and the experimental results are shown in Fig. 3

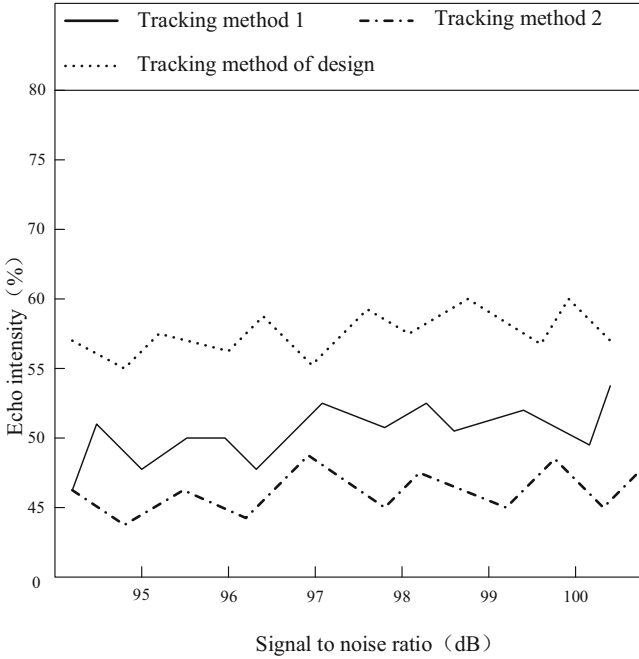


(a) 10 dB SNR

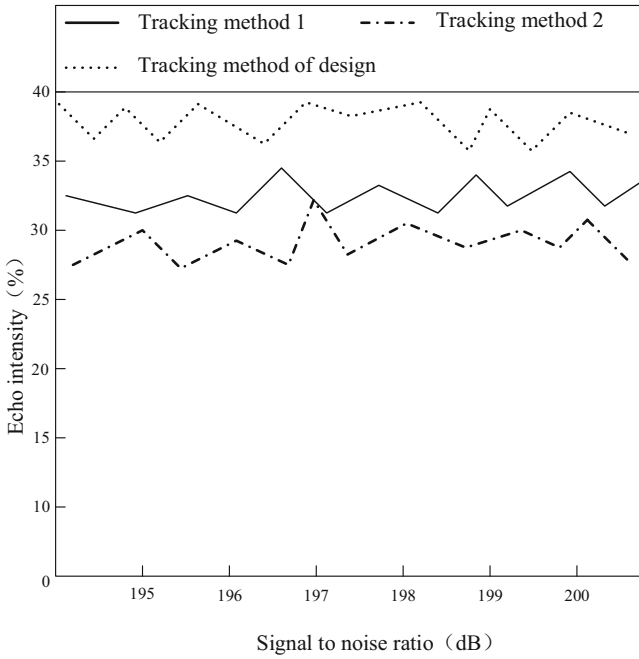


(b) 50 dB SNR

**Fig. 3.** Experimental results of echo intensity

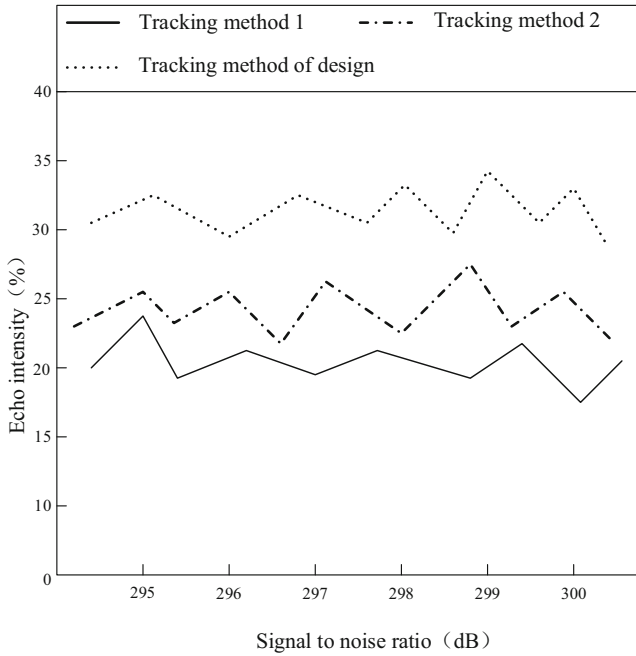


(c) 100 dB SNR

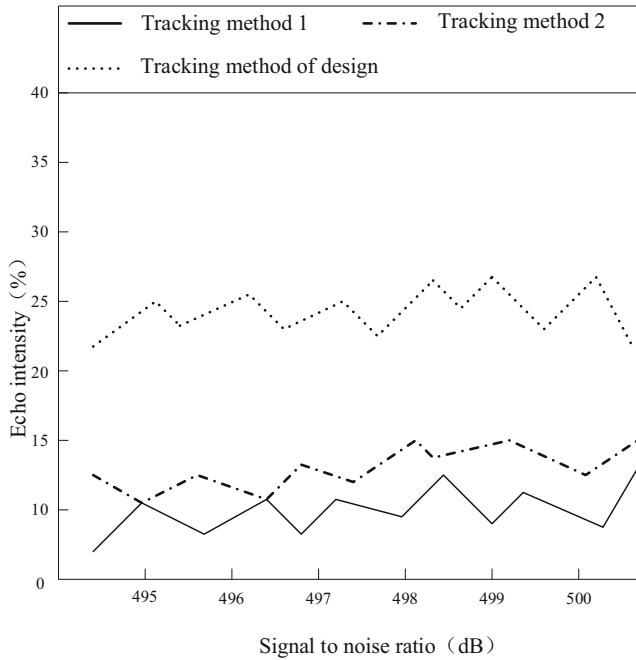


(d) 200 dB SNR

Fig. 3. (continued)



(e) 300 dB SNR



(f) 500 dB SNR

Fig. 3. (continued)

It can be seen from Fig. 3 that under different SNR conditions, the average echo intensity of the designed tracking method and the two existing target tracking methods are 56.464%, 39.887% and 41.203% respectively, which proves that the designed target tracking method is more practical.

## 4 Conclusion

Wake radar detection technology is generally used to enhance the flight safety of civil aviation aircraft and improve transportation efficiency. In the field of military applications, because the electromagnetic scattering characteristics of aircraft wake are closely related to aircraft wingspan, takeoff weight and flight speed, the wake radar detection technology is expected to be applied to aircraft target detection, recognition and other fields. Therefore, this paper proposes an X-band radar detection target tracking method based on IOT sensing technology, and carries out experimental research. The target tracking method designed in this paper is proved to have better performance than the existing tracking methods by experimental test, which enriches the academic literature on target tracking. At the same time, it broadens the application field of IOT perception technology to a certain extent, and lays a theoretical and practical foundation for academic research. Due to the limited research conditions, the research on the application of IOT sensing technology in other fields is not thorough enough, and will continue to improve in the future.

## References

1. Chen, X., Liu, Y., Hong, X., et al.: Unmanned boat target detection and tracking method based on SSD-CF. *China Measurement Testing Technol.* **45**(2), 145–150 (2019)
2. Tian, J., Jiang, F.: Laser detection and control system based on Internet of things technology. *Laser J.* **40**(3), 153–157 (2019)
3. Ji, L., Zhang, Y., Wang, Y., et al.: Research on the spectral width of the X-band radar clutter backscattered from sea surface. *Acta Oceanologica Sinica* **41**(7), 149–158 (2019)
4. Qinglong, B.A.O., Sen, W.A.N.G., Jiameng, P.A.N., et al.: Key technology analysis of target detection system based on non-cooperative radar illuminator. *Chin. J. Radio. Sci.* **35**(4), 496–503 (2020)
5. Zhu, G., Zhang, Z.: An adaptive beam resource allocation strategy of networked radar for target tracking. *Electron. Optics Control* **27**(3), 75–79 (2020)
6. Chai, J., Du, W., Yuan, Q., et al.: Analysis of test method for physical model test of mining based on optical fiber sensing technology detection. *Optical Fiber Technol.* **48**, 84–94 (2019)
7. Qi-song, Y.A.N.G., Yong-gang, J.I., Yong-xin, L.I.U., et al.: Target tracking of surface wave over-the-horizon radars based on FJPDA algorithm. *Adv. Marine Sci.* **37**(1), 84–91 (2019)
8. Li, S., Icheng, T.: Interactive multiple model algorithm for a doppler radar maneuvering target tracking based on converted measurements. *Acta Electronica Sinica* **47**(3), 538–544 (2019)
9. Xiong, J., Zuo, Z., Xiong, J.: An improved airborne doppler radar target tracking method. *Telecommun. Eng.* **59**(9), 1026–1030 (2019)



10. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
11. Zheng, S., Wang, Z., Xu, L.: Research on the application of photoelectric detection information in air defense missile weapon system. *Aerospace Control* **37**(2), 15–18 (2019)
12. Liu, S., Liu, X., Wang, S., Muhammad, K.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
13. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing*, (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>
14. Li, X., Yang, Y., Sun, Z., et al.: Multi-frame integration method for radar detection of weak moving target. *IEEE Trans. Vehic. Technol.* **PP**(99), 1 (2021)
15. Rapid wavelet-based bathymetry inversion method for nearshore X-band radars. *Remote Sensing of Environment*, **240**, 111688 (2020)
16. Zhao, C., Zhang, Y., Zheng, D., et al.: An improved hydrometeor identification method for X-band dual-polarization radar and its application for one summer Hailstorm over Northern China. *Atmos. Res.* **245**, 105075 (2020)
17. Jianli, M.A., Zhiqun, et al.: Improvement of X-band polarization radar melting layer recognition by the bayesian method and ITS impact on hydrometeor classification. *Adv. Atmospheric Sci.* **37**(01), 107–118 (2020)
18. Zhang, M., Liu, X., Xu, D., et al.: Vision-based target-following guider for mobile robot. *IEEE Trans. Ind. Electron.* **PP**(99), 1 (2019)
19. Maddox, S.J., Dunne, L.: MADX – a simple technique for source detection and measurement using multi-band imaging from the Herschel-ATLAS survey. *Mon. Not. R. Astron. Soc.* **2**, 2 (2020)
20. Zhou, Y., Wang, T., Hu, R., et al.: Multiple Kernelized Correlation Filters (MKCF) for extended object tracking using X-band marine radar data. *IEEE Trans. Signal Process.* **67** (14), 3676–3688 (2019)
21. Paranthaman, R., Moses, J.A., Anandharamakrishnan, C.: Development of a method for qualitative detection of lead chromate adulteration in turmeric powder using X-ray powder diffraction. *Food Control* **2021**(7), 107992 (2021)
22. Li, P., Xie, S., Wang, K., et al.: A novel frequency-band-selecting pulsed eddy current testing method for the detection of a certain depth range of defects. *NDT and E Int.* **111**, 102154 (2019)
23. Yu, Y., Wang, H., Liu, S., et al.: Distributed multi-agent target tracking: a nash-combined adaptive differential evolution method for UAV systems. *IEEE Trans. Vehicular Technol.* **PP**(99), 1 (2021)



# Research on Hybrid Encryption of Cross Border E-commerce Transaction Information Based on B+ Search Tree Algorithm

Jia-hua Li<sup>(✉)</sup>

School of Information Engineering, Guangzhou Vocational and Technical University of Science and Technology, Guangzhou 510550, China  
lijiahua21223@163.com

**Abstract.** In order to solve the problem of low security existing in the traditional information hybrid encryption method, a cross border e-commerce transaction information hybrid encryption method based on B+ search tree algorithm is proposed. In the cross border e-commerce trading platform, B+ search tree algorithm is used to retrieve the e-commerce trading data, process and cluster the e-commerce trading information. On this basis, idea and ECC are taken as the hybrid encryption algorithm types, and the hybrid encryption of cross border e-commerce trading information is realized through the steps of generating key and private key, key management and distribution, and key exchange in hybrid encryption. Compared with the traditional encryption method, the experimental results show that the decryption time of the hybrid encryption method designed in this paper is longer, and the amount of data error is reduced by 2.44 MB, which means that the security of the method is higher.

**Keywords:** B+ search tree algorithm · Cross border transactions · E-commerce transaction information · Hybrid encryption

## 1 Introduction

E-commerce is a new type of business activity which is emerging in the current economic and social state. E-commerce has two types of explanation, narrow sense and broad sense. In a narrow sense, e-commerce refers to the activities of commodity transaction between consumers and consumers, between consumers and enterprises, enterprises and enterprises through the Internet. In a broad sense, e-commerce refers to all the comprehensive e-commerce activities carried out through the Internet/ Intranet/enterprise external network [1]. In recent years, with the further acceleration of globalization, e-commerce, as a product of market economy and financial industry development and integration, has been developed rapidly. According to the current transaction scale, the future global trade economy will be traded through the network in the next ten years [2]. Under the trend of global economic integration, e-commerce will be able to trade in the future, In the future, e-commerce will occupy a greater and larger part in economic and trade. With the development of computer network technology and information technology, human beings have entered the information society and

network economy era rapidly. The development of e-commerce is becoming more and more rapid [3]. According to the annual monitoring report of China's e-commerce industry in 2019–2020, the e-commerce market in China will maintain a high-speed development in 2020, with the transaction scale exceeding 18.1 trillion yuan, In 2020, the transaction volume of China online shopping market amounted to 3304.53 billion yuan, accounting for 13.2% of the total retail sales of social consumer goods, with the penetration scale of online shopping users reaching 56.9%. Online transaction has become a mainstream consumer category. The trading platform emerges like a springing up. Nowadays, the mainstream trading platforms include tmall, Taobao, JD, etc., and the trading scale is expanding [4]. Therefore, a large amount of data is generated, and the security of cross border e-commerce transaction information is also seriously threatened. Therefore, the encryption method of cross border e-commerce transaction information is proposed by using cryptography theory.

The basic idea of cryptography is to camouflage confidential information. A cryptosystem completes the following disguise: a user (encryptor) transforms (encryption transformation) the confidential information (clear text) that needs to be disguised to obtain another representation (ciphertext) which seems to be irrelevant to the original information [5]. If the legitimate user (receiver) obtains the disguised information, he can restore the original confidential information (decryption transformation) from the information, if illegal users (password analysts) try to get the original confidential information from the disguised information, either such analysis is impossible at all, or the cost is too large to be carried out. At present, the mature encryption methods include DES encryption algorithm, IDEA encryption algorithm, ECC encryption algorithm, RSA encryption algorithm, etc. from the current domestic and foreign research status, it is necessary to achieve the ideal goal of high efficiency and security in information transmission, and the development of encryption technology to the cross use of RSA and AES is inevitable. However, different encryption algorithms have different defects, such as RSA has a slow encryption speed and low security of AES encryption results.

At present, reference [6] proposed a data encryption method based on adversarial neural network. By analyzing the idea of counterattack neural network and introducing the counterattack game theory of cryptography, a counterattack neural network encryption method based on selective ciphertext attack is proposed. By introducing the attacker of selective ciphertext attack, the existing encryption algorithm of neural network is improved to complete the data encryption processing. Reference [7] proposed a data encryption method based on cloud computing. This method firstly analyzed the traditional symmetric encryption algorithms DES and AES, and compared their advantages and disadvantages in the cloud computing model. Then, combined with Hadoop distributed computing framework, an improved DAES hybrid encryption algorithm is proposed. This algorithm can effectively improve the security by mixed encryption of plaintext partition and adding random disturbance information.

In order to solve the problems existing in traditional encryption algorithm, a hybrid encryption method is proposed. The principle of this method combines source encryption with line encryption. The specific method is: users encrypt plaintext, and encrypted ciphertext directly enter the line secret machine, perform a second encryption. In order to improve the mixed encryption effect of cross border e-commerce

information and ensure the security of cross border e-commerce transaction information, B+ search tree algorithm is introduced. B+ tree is a variant tree of B-tree which is required by file system. Usually, there are two head pointers on B+ tree, one pointing to root node and one pointing to leaf node with the smallest keyword. With the application of B+ search tree algorithm, the operation security of cross border e-commerce transaction information is improved to the lowest extent by mixing two or more encryption algorithms.

## 2 Design of Hybrid Encryption Method for Cross Border E-commerce Transaction Information

### 2.1 Using B+ Search Tree Algorithm to Retrieve E-commerce Transaction Data

In the basic B-tree, the keywords are distributed in the whole B-tree, and the keywords that appear in the internal nodes no longer appear in the leaf nodes, so that the order chain cannot connect all the keywords in the tree. B+ tree has changed this point, that is, all keywords in the tree are inserted on the leaf node from left to right in increasing order, and connected with pointers. In B+ tree, data pointers are only stored in the leaf node of the tree. Therefore, the structure of leaf node is different from that of internal node. If the search field is a keyword, the leaf node has an entry and a pointer to the record for the value of each search field. For non key search fields, the pointer points to a block in the add-on level, which stores the record pointer to the data file [8]. The structure of B+ search tree is shown in Fig. 1.

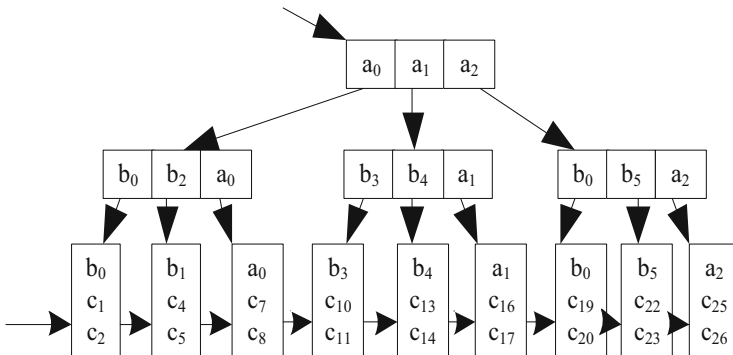


Fig. 1. B+ search tree structure

It can be seen from Fig. 1 that B+ search tree is a three-layer structure from top to bottom, and all the key codes appear on leaf nodes. The key codes in nodes of each layer above are copies of the largest key codes in corresponding nodes of the next layer. Since the construction of B+ tree is bottom-up, M limits the size of nodes and

copies the maximum key code of each node to the node of the previous layer from bottom to top.

In order to enable searchers to quickly find the transaction identifiers of ciphertext documents related to trap door transactions on the block, the data owner stores the transaction identifiers related to ciphertext transactions in the leaf nodes of B+ tree, where the structure of the nodes is shown in formula 1.

$$u = (N, D, P_1, P_r, TXid) \tag{1}$$

where,  $N$  is the node number in the tree index structure,  $D$  is the  $m$  dimensional vector composed of  $m$  keywords of each document,  $P_1$  and  $P_r$  are the left and right pointers of node  $u$  respectively. If  $u$  is the leaf node of the index tree, the transaction identifier txid of the ciphertext document is stored in the leaf node, and if  $u$  is the internal node, the calculation of node vector  $D$  is shown in Eq. 2.

$$D[j] = \max\{u.P_1 \rightarrow D[j], u.P_r \rightarrow D[j]\} \tag{2}$$

The process of inserting nodes into B+ tree is the process of building ciphertext transaction index. In the actual cross border e-commerce transaction information retrieval process, we need to search the B+ tree node and find keywords in the B+ node on the cross border e-commerce platform to get the corresponding transaction information retrieval results [9]. Suppose that the B+ tree of order  $m$  contains  $N$  keywords, and the analysis shows that there are at least  $2 \times \left[ n^{\frac{2}{m}} \right]^{k-1}$  nodes in layer  $k$ . If the B+ tree is inserted, it will split. Let  $L$  be the number of internal nodes. When all internal nodes except the root contain  $\left[ n^{\frac{2}{m}} \right]^{-1}$  keywords, the B+ tree contains the least number of total keywords:

$$N \geq 1 + L \left( n^{\frac{2}{m}} - 1 \right) \tag{3}$$

That is to say, in the worst case, each inserted node of  $L$  is split, that is,  $L$  is split, then the average number of split nodes for each inserted key is:

$$S = \frac{K}{N} \tag{4}$$

In the B+ search tree algorithm, through the cross border e-commerce trading platform of each node location data search, get the final search results.

## 2.2 Data Acquisition and Processing of E-commerce Transaction Information

In the process of data collection, the first step is to simulate the operation of the browser on the page, and then analyze the structure of the page through the given URL. For ordinary pages, the basic Http Client method is used to obtain them. For special pages,

it is necessary to further determine whether they are dynamically loaded pages or pages that need further interaction to obtain data [10]. For pages with dynamic secondary loading information, Htmlunit provides `Web Client.get Options().set Java Script Enabled()` method to parse JavaScript scripts. For Ajax, the `web client. Set Ajax controller ()` method is also provided to support Ajax. Therefore, htmlunit can directly parse the dynamic secondary loading information page. The transaction statement set can be obtained from the transaction web page data after preprocessing, and these transaction statement sets still contain a lot of noise, such as non transaction statement, invalid transaction, etc. if the transaction statement is analyzed and extracted directly, the results may deviate from the actual situation, or even draw wrong conclusions. Therefore, we need to filter and clean the product transaction statement set again. In the actual processing process, the noise in the product transaction sentence set is mainly manifested in the following aspects: the non transaction noise such as characters irrelevant to the transaction sentence and explanatory text, which will increase the time consumption of further analysis of the transaction sentence, and also cause interference to the syntactic analysis and opinion extraction of the transaction sentence; Automatic transactions of e-commerce platform, such as “default praise” transactions, have no significance for the extraction and analysis of the whole transaction view; In order to obtain benefits, some businesses or competitors mislead consumers by publishing statements to promote or slander a certain brand product, or even asking the Internet water army to publish false transactions, which will cause interference to the analysis results of the overall view of the transaction and need to be filtered out; For non trading statements in Product Trading Web data, direct filtering method can be adopted to extract trading statements. For noise trading in trading statement set, for unrelated product trading statements and non trading statements, firstly, the garbage trading set is obtained by manually marking training set, and then the machine learning model is established by logistic regression to identify these two types of invalid transactions; For the deceptive trading information such as defamation, promotion and interference, the repetitive trading information is identified by machine learning model, which takes the repetitive trading data statements as the positive training set.

### **2.3 Clustering Cross Border E-commerce Transaction Information**

Clustering is to divide a physical or abstract data set into several classes. The data in each class has great similarity, and there is little similarity or dissimilarity between classes. Its analysis basis is the similarity calculation between data. By clustering the cross border e-commerce transaction information, it is convenient for the hybrid encryption algorithm to process the same type of data, so as to improve the encryption effect. The clustering process of cross border e-commerce transaction information is shown in Fig. 2.

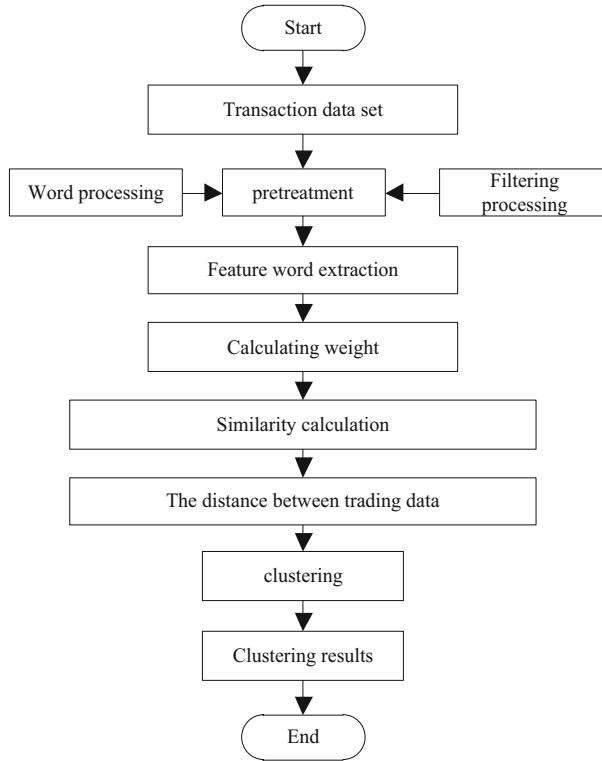


Fig. 2. Cross border e-commerce transaction information clustering flow chart

In clustering analysis of product transaction data statements, it is necessary to calculate the similarity of transaction data statements, and classify the statements with similar transaction contents into one category. When the transaction data is vectorized, in order to reduce the dimension of the constructed sentence vector, all the words in the sentence can not be expressed in the vector. It is necessary to extract the words that can represent the content and attributes of the transaction data in the transaction sentence, that is, the feature words, and extract them to form the vector representation of the transaction sentence [11]. Transaction data vectorization is to abstract each feature word of transaction data into a component of sentence vector. For the convenience of description, the transaction data to be classified is defined as:

$$D = \{S_k, K = 1, 2, 3 \dots\} \tag{5}$$

After  $S_k$  is processed by word segmentation, pronoun removal and auxiliary word removal, the remaining words are used as feature words to construct the vector model of the sentence, and the set of feature words of  $S_k$  is represented as  $S_{kw}$ , which is the basis of constructing the similarity calculation model. After the feature word set representation of the sentence is obtained, the weight of each feature word in the set in the

transaction data is calculated, and the weight value of each feature word is composed of the following vectors:

$$V_s = (v_1, v_2, \dots, v_k) \quad (6)$$

$v_i$  represents the weight of the corresponding feature word  $w_i$  in the transaction data, which constitutes a  $k$ -dimensional space vector and realizes the vectorization of the transaction data.

In order to meet the requirements of processing time when processing text big data, the weight value of each feature word in the transaction data vector is calculated [12]. Usually, the importance of a feature word increases with the frequency of its appearance in the document, but decreases with the frequency of its appearance in the corpus. The word frequency TF represents the frequency of feature word  $w_i$  appearing in transaction data  $D$ , and the importance of  $w_i$  to document  $D$  is expressed as follows:

$$TF = \frac{N_{w_i}}{N_{Dw_i}} \quad (7)$$

where  $N_{w_i}$  is the number of times that feature word  $w_i$  appears in document  $D$ , and  $N_{Dw_i}$  is the total number of words in document  $D$ . In addition, the inverse document frequency *IDF* represents the number of documents containing the feature word  $w_i$  in the corpus, and the value of *IDF* is calculated by using the following formula:

$$IDF = \lg \frac{N_D}{N_{w_i D}} \quad (8)$$

where  $N_{w_i}$  is the number of times that feature word  $w_i$  appears in document  $D$ , and  $N_{Dw_i}$  is the total number of words in document  $D$ . In addition, the inverse document frequency *IDF* represents the number of documents containing the feature word  $w_i$  in the corpus, and the value of *IDF* is calculated by using the following formula:

$$v_i = TF \times IDF \quad (9)$$

In order to improve the efficiency of feature word weight calculation, the map reduce framework is used to make word statistics on the transaction statements of each storage node, and the statistical results are summarized. When calculating TF and IDF values in TF-IDF algorithm, there is no need to repeat the statistical operation, which saves the statistical time and improves the efficiency of the algorithm. In the cross border e-commerce transaction information, select the appropriate clustering center, and calculate the similarity of transaction information [13]. The weight of the feature words of two sentences in the transaction data is used to form the vector representation of the sentence, that is, the feature vector of the sentence is obtained, and then the cosine value of the angle between the feature vectors is used as the similarity value of the sentence:



$$Sim(S_1, S_2) = \frac{\sum_{i=1}^n \phi_i \cdot \psi_i}{\sqrt{\sum_{i=1}^n \phi_i^2} \cdot \sqrt{\sum_{i=1}^n \psi_i^2}} \tag{10}$$

$\phi_i$  and  $\psi_i$  represent the eigenvector components of sentence  $S_1$  and  $S_2$  respectively. Based on the above operations, the number of cluster centers  $K$  is initialized by K-means algorithm, and  $K$  transaction statements are randomly selected as cluster centers. The distance between each transaction information and  $K$  cluster centers is calculated respectively, and each transaction is allocated to the cluster with the nearest cluster center, and the cluster center of each family is recalculated. Finally, determine whether the new cluster center and the original cluster center are less than the threshold value. If it is less than the threshold, the cluster results will be output directly, otherwise, the above steps need to be performed again. The final result of clustering processing of cross border e-commerce transaction information is shown in Fig. 3.

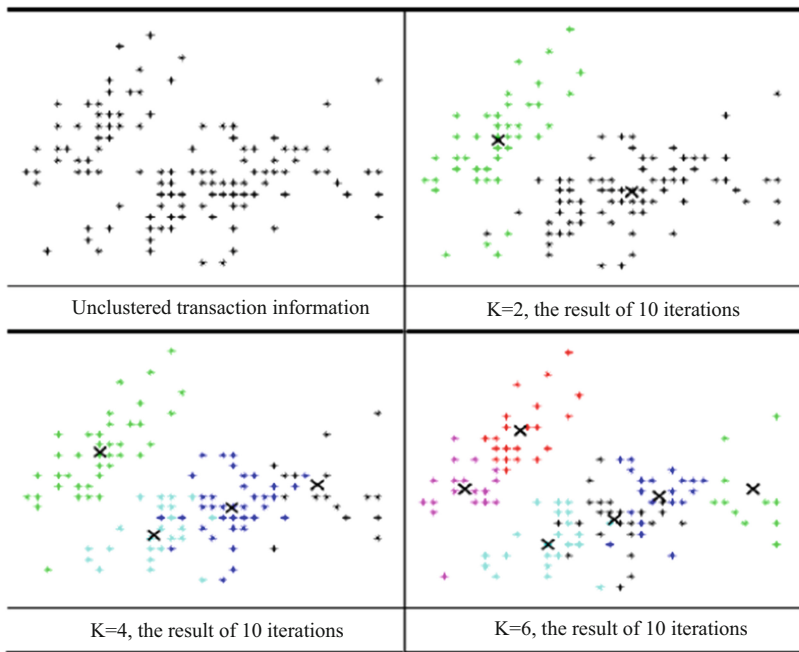


Fig. 3. Clustering effect of e-commerce transaction data under different K values

### 2.4 Select Hybrid Encryption Algorithm Type

In accordance with the principles of accuracy, simplicity, efficiency and practicability, the type of hybrid encryption algorithm is selected. This paper compares the

performance of encryption algorithm from two aspects of security and encryption speed, and the comparison results of encryption algorithm performance are shown in Table 1.

**Table 1.** Performance comparison of encryption algorithms

Encryption algorithm type	How long does it take to decode the key/h	Key length/bit
AES	108	768
DES	104	512
RSA	204	768
MD5	1011	21000
IDEA	1020	2048
ECC	1078	1024

As can be seen from Table 1, compared with other asymmetric cryptography technologies, IDEA and ECC cryptography technologies have absolute advantages in anti attack and cracking performance. Through continuous testing, scientists have come to the conclusion that if IDEA and ECC want to decode, the difficulty is far greater than that of algorithms such as RSA, and they are not on the same level at all. If ECC and RSA are in the same computing conditions, RSA algorithm can reduce the length of public key to improve the encryption efficiency, and improve the speed of digital signature and authentication. However, in the private key processing level, elliptic curve algorithm has much higher data processing efficiency than the other two algorithms. And its key generation speed is much faster than the other two. Therefore, if these algorithms are in a peer-to-peer computing environment, the overall resource consumption and processing efficiency of ECC algorithm are much better. To sum up, IDEA encryption algorithm and ECC encryption algorithm are selected as hybrid algorithms.

IDEA algorithm is a symmetric key block cipher system, its plaintext and ciphertext are 64 bits, and the key length is 128 bits. The data is encrypted in 64bits packets. The encryption process is to iterate the input 64bits plaintext for 8 rounds, and the resulting iteration result is transformed into the final output 64bits ciphertext group. ECC encryption algorithm is also known as elliptic curve encryption algorithm. The order of the curve is defined as the number of all points on the elliptic curve, denoted as  $E$ , which satisfies the following relationship:

$$p + 1 - 2\sqrt{p} \leq E \leq p + 1 + 2\sqrt{p} \quad (11)$$

The finiteness of points and the uncertainty of the number of points on elliptic curve group are good attributes for encryption, because these curves only contain some discrete points, and the attacker does not know how to apply the geometric relationship. Suppose that a point  $P$  on an elliptic curve satisfies the minimum positive integer  $n$  of number multiplied by  $n_P = 0$ , which is called the order of point  $P$ . The implementation of elliptic curve cryptosystem is determined by the domain parameters of elliptic curve, which include the base domain of elliptic curve, curve equation, base

point of curve and order of base point. The parameters of elliptic curve domain are public and shared by both sides of communication. Randomly select an integer  $d$  in the interval  $[1, n - 1]$  and calculate the formula as follows:

$$Q = dP \tag{12}$$

where  $d$  is the private key and  $Q$  is the public key. The following nodes are specific to the body area network application, node  $A$  represents the collection node, and node  $B$  represents the sink node. In the encryption process, when node  $A$  sends physiological information to node  $B$ , node  $A$  encodes the physiological information  $M$  to be transmitted to a point  $m$  on  $E_p(a, b)$ . Node  $A$  looks up node  $B$ 's public key  $Q_B$  and public information base point  $P$ , and transmits encrypted data  $C_1$  and  $C_2$  to node  $B$ .

### 2.5 Realize Cross Border E-commerce Transaction Information Hybrid Encryption

Taking node  $B$  as the collection node to send physiological information to sink node  $A$  as an example, in the previous discussion, the key encapsulation mechanism will include three sub processes, namely, the generation of  $A$  and  $B$  public keys, the key encapsulation of  $B$  and the key unsealing of  $A$ . The core idea of the algorithm is that  $A$  and  $B$  use the same shared secret key. The specific idea is that  $A$  uses its own private key and  $B$ 's public key to carry out the corresponding operation. At the same time, node  $B$  uses  $B$ 's private key and  $A$ 's public key to carry out the same operation. On the basis of elliptic curve encryption, the key encapsulation of hybrid encryption system is shown in the figure below (Fig. 4).

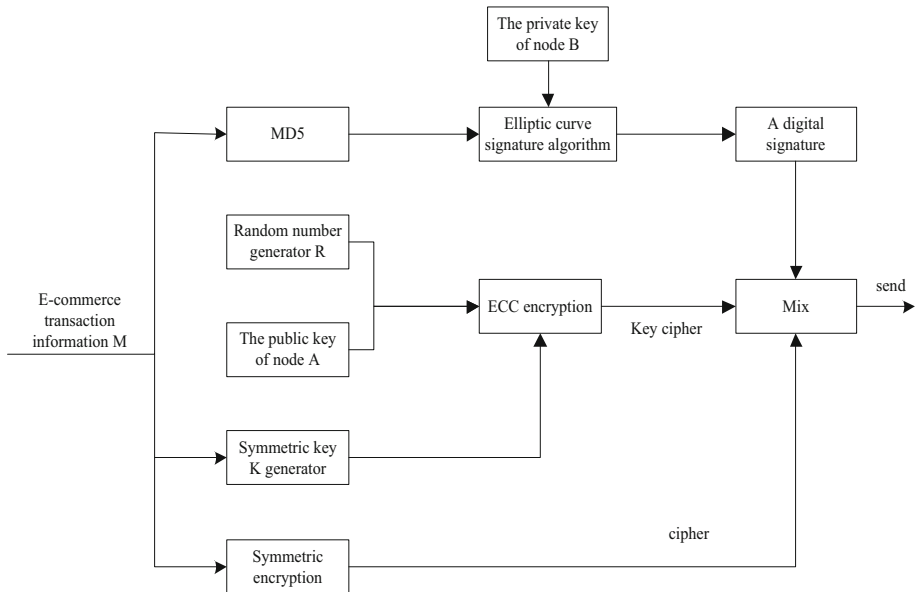


Fig. 4. Block diagram of hybrid encryption key encapsulation

### Generate Key and Private Key

Through the combination of idea and ECC algorithm, the following mixed encryption expression is formed

$$\begin{cases} x_{n+1} = ax_n(1 - x_n) \\ y_{n+1} = by_n(1 - y_n) \end{cases} \quad (13)$$

where,  $x_n$  and  $y_n$  are the input items of cross border e-commerce transaction information,  $x_{n+1}$  and  $y_{n+1}$  are the final encryption results, and  $a$  and  $b$  are constant coefficients. In the process of generating ECC private key, the first step is to input the initial value  $(X_0, Y_0)$ , the number of iterations  $n$ , and the parameters  $a$  and  $b$ ; The input parameters are used as the input parameters of the hybrid encryption method for iteration until the required number of iterations  $n$  is reached, and the final iteration result is returned. Calculate the result of  $X \times Y$  is denoted as  $T_1$ , and the integer of  $T_1$  is denoted as  $T$ ; If  $T$  is less than the multiple of ECC's base point, then  $T$  is assigned to key as ECC's private key; Otherwise, return to the first step.

### Key Management and Distribution

The key distribution includes the distribution of public key and secret key. The public key distribution is also the problem of how to obtain the public key of the other party in the system of public key cryptosystem. First, the key and public key are obtained by using the key generation software. The private key is in the local file and cannot be passed out. As for public key is public password, it can be known to others. Then the central computer needs to save the KU receiving of each node in the local special file system. In this component, the data encryption transmission process is: in the initial stage, the background center computer generates an encryption key. If it is called DEKEY, then a key for signature is generated. If it is called MKEY, both keys need to be saved. Then, we need to encrypt two keys by using elliptic curve algorithm, and the encrypted key is. Then, the encrypted DEKEY and MKEY are transmitted to the front desk. The client decrypts the data to get key and MKEY by using the private DEKEY generated by elliptic curve after receiving the file, and saves the obtained file in the local system.

### Key Exchange in Hybrid Encryption

The key exchange method with the ability of confidentiality and authentication is adopted. The method is assumed to be carried out when the public key of both parties is consistent with each other. In fact, the certification authority now guarantees the completion of this part of the work. The confidentiality and identification are realized through two repeated verification of  $A$  and  $B$  parties. Because of the universality of encryption of common key algorithm key with public key algorithm, the scheme has a wide range of applications.  $A$  encrypts a message sent to  $B$  with the public key of  $B$ , which contains a random identifier  $ID_A$  generated by  $A$ , and  $B$  sends a message encrypted by mixed encryption method to  $A$ , which contains identifier  $ID_B$  generated by identifier  $ID_A$  and  $B$  of  $A$ .  $A$  returns a  $ID_B$  encrypted with the public key of  $B$  to

make  $B$  sure the other party is  $A$ ,  $A$  selects a secret key  $K$  and sends the program parameter represented by formula 14 to  $B$ .

$$M = (E_{KP_b}(E_{KP_a}(K))) \tag{14}$$

where  $(E_{KP_a}, E_{KP_b})$  generates a public and private key pair for  $A$ . Encryption with  $B$ 's public key ensures that only  $B$  can interpret it; Encryption with the private key of  $A$  ensures that only  $A$  can send it. After receiving it,  $B$  side recovers the secret key  $K$ , and realizes confidentiality and authentication through two repeated verifications of  $A$  and  $B$ .

**Cross Border E-commerce Transaction Information Mixed Encryption Transmission**

The hybrid cryptographic communication process mainly involves the operation of the sender and the receiver, in which the sender's operation is regarded as the encryption process and the receiver's operation as the decryption process. The generated key and plaintext are digitally encapsulated, and the information is transmitted through the communication transmission channel. After receiving the data, the receiver can decrypt the key and ciphertext through digital signature verification.

**3 Experimental Analysis of Encryption Performance Test**

In order to test the encryption performance of the hybrid encryption method of cross-border e-commerce transaction information based on B+ search tree algorithm, the performance test experiment is designed, and the operational advantages of the design method are reflected through quantitative comparison with different methods.

**3.1 Development Tool and Test Environment of Encryption Method**

The experiment of hybrid encryption algorithm is carried out on the laboratory Hadoop experimental platform. The experimental platform consists of six computers in the laboratory. According to the structure requirements of Hadoop platform, one of the six computers acts as the namenode server, which is responsible for the scheduling control of the whole system. The other five computers are used as storage nodes and computing nodes. Due to the limitation of computers owned by the laboratory, the configuration of these five computers is not the same, and the purchase age is also different. In addition, during the experiment, these five computers were added to Hadoop cloud computing platform. Now the configuration of these six computers is briefly introduced, as shown in Table 2.

**Table 2.** Configuration of computers used in the experiment

Name	CPU	Frequency	Memory	Hard disk
Cloud2	Pentium 4	3.0G	512 MB	80G
Cloud3	Pentium 4	3.0G	512 MB	80G
Cloud4	Pentium 4	3.0G	512 MB	80G
Node1	IS-2320	3.0G	3G	150G
Node2	IS-2320	3.0G	3G	150G
M290	IS-2320	3.0G	3G	150G

It can be seen from Table 2 that among the six computers, the computing power and storage capacity of three computers are better than those of the other three computers. In this case, according to the experimental requirements, the operator selects the computer with higher computing power and storage capacity as the computing node and storage node. This design also meets the requirements of high computing power of hybrid encryption algorithm experiment. In addition to the hardware requirements, the experimental platform also has certain software requirements. The Hadoop cloud computing platform composed of six computers in the laboratory uses Ubuntu 10.10 as the operating system and Hadoop 0.20.203 as the Hadoop system version. In addition, because the processing object of the encryption method exists in the form of document, and the designed encryption method applies B+ search tree algorithm, it is necessary to embed multiple processing programs in the experimental environment. For example, Weblogic 10 is used as the development server. The WebLogic Server configures the JSP container by configuring the weblogic.xml file. The weblogic.xml file can be used in the web of the configured web application\_Inf directory. Generally speaking, it is not necessary to modify the Weblogic. XML file manually for developing and deploying web applications. When deploying the application again, you can configure it in the console. Select the deployment option in the console and click webapplication to configure the web application. Then the corresponding configuration file can be generated in the system. Developers can configure different information in development and production environments.

### 3.2 Prepare Cross Border E-commerce Transaction Data Samples

The cross border e-commerce transaction data encrypted in the experiment mainly comes from the cross border e-commerce transaction platform. In this experiment, the home decoration design e-commerce platform is selected to provide transaction data samples for the experiment. The operation of the platform is shown in Fig. 5.

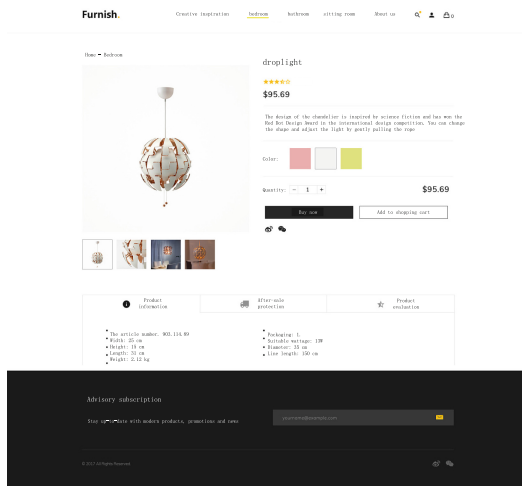


Fig. 5. Cross border e-commerce transaction interface

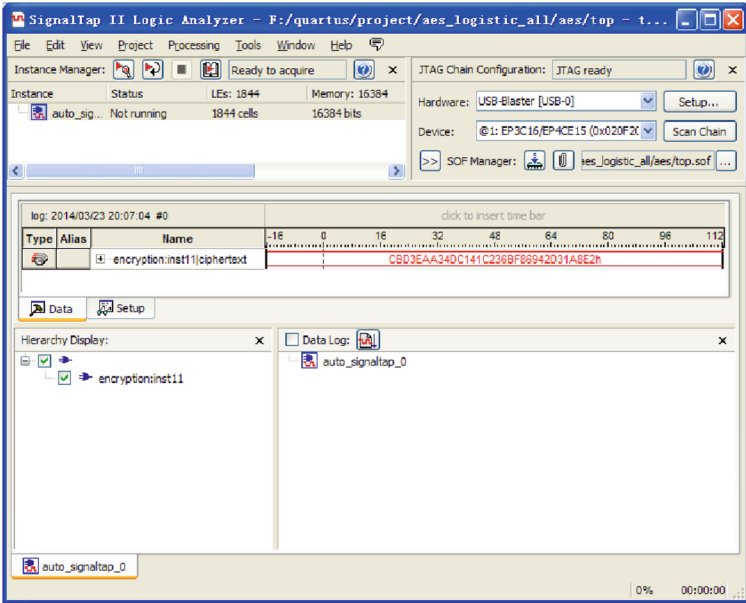
All transaction data in the trading platform are collected as data samples during the half year from October 2020 to March 2021, with a total of 24.17 GB of transaction data.

### **3.3 Set up Experimental Comparison Items and Test Indexes**

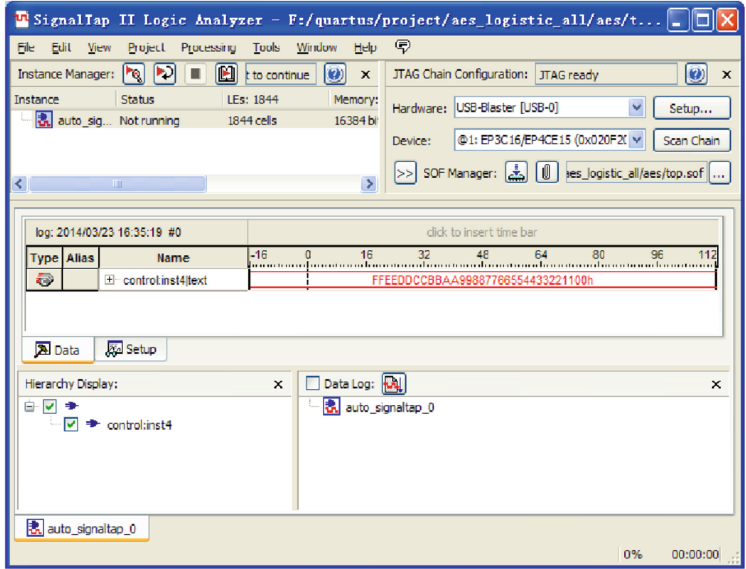
In order to reflect the performance of the design encryption method, the encryption method based on the adversarial neural network and the encryption method based on cloud computing are set as the two contrast methods of the experiment, and the two contrast methods are imported into the experimental development and operation environment in the same way. In the experiment, the cross-border e-commerce transaction data processed by the three encryption methods are the same, so as to ensure the uniqueness of the experimental variable. The security of the encryption result is set as the test index of the encryption performance, which consists of two parts, namely the key cracking time and the data error amount of the decryption result. The longer the key cracking time, the higher the encryption effect, while the more the data error amount of the decryption result, the lower the security of the encryption method.

### **3.4 Describe the Experimental Process of Encryption Performance Test**

In order to effectively simulate the actual cross border e-commerce trading platform environment, the attack program is used to attack the transaction data, and the attack frequency and intensity are controlled. According to the attack settings, the experimental groups are divided, and the final comprehensive test results are obtained by averaging multiple experiments, so as to enhance the credibility of the experimental conclusions. Through the operation of the three encryption methods, the corresponding encryption results are obtained. The output results of the designed hybrid encryption method are shown in Fig. 6.



(a)Ciphertext results obtained by hybrid encryption



(b)Plaintext sequence received by receiver

Fig. 6. Cross border e-commerce transaction information mixed encryption results



### 3.5 Analysis of Test Results

Through the statistics and analysis of relevant data, the quantitative test results of key cracking time are obtained, as shown in Table 3.

**Table 3.** Statistics of key cracking time test(h)

Experimental group	Encryption method based on adversarial neural network	Encryption method based on cloud computing	Design encryption method
1	16.4	25.4	35.4
2	19.8	22.8	42.6
3	21.4	26.5	38.9
4	18.6	27.1	34.1
5	20.4	23.5	33.7
6	17.5	22.7	45.1
7	14.3	24.4	37.4
8	18.4	26.3	42.2

It can be seen from Table 3 that the average key cracking time of the three encryption methods is 18.35 h, 24.84 h and 38.68 h respectively, which shows that the key cracking time of the encryption method is longer.

In addition, the results of the three encryption methods are decrypted, and the statistical results of the amount of data errors are obtained under different attack environments, as shown in Table 4:

**Table 4.** Statistics of decryption data error

Experimental group	Encryption method based on adversarial neural network		Encryption method based on cloud computing		Design encryption method	
	Data loss/MB	Amount of error data/MB	Data loss/MB	Amount of error data/MB	Data loss/MB	Amount of error data/MB
1	1.14	1.25	0.56	0.56	0.14	0.04
2	1.28	1.31	0.52	0.48	0.07	0.05
3	1.06	1.45	0.78	0.42	0.08	0.02
4	1.26	1.52	0.69	0.63	0.13	0.04
5	1.31	1.40	0.61	0.48	0.11	0.11
6	1.36	1.26	0.72	0.44	0.05	0.06
7	1.29	1.31	0.59	0.61	0.06	0.03
8	1.44	1.09	0.49	0.75	0.10	0.07

Through the statistics of the data in Table 4, we can see that the data errors of the three encryption methods are 2.59 MB, 1.17 MB and 0.15 MB respectively. To sum up, from the two aspects of the key cracking time and the amount of decrypted data errors, the security of the design encryption method is higher.

## 4 Concluding Remarks

As a new data encryption algorithm, hybrid encryption algorithm will be more effective, secure and flexible than the past simple encryption/decryption algorithm. The hybrid encryption method combining IDEA and ECC not only solves the difficulty of key distribution, but also solves the speed and efficiency of encryption and decryption. Undoubtedly, it is a better and feasible method to solve the problem of information security. According to the diversity of existing encryption algorithms, hybrid encryption/decryption algorithm has a better development prospect.

**Fund Projects.** 1. Key scientific research platforms and projects of ordinary universities in Guangdong Province in 2020-Research on cross-border E-commerce blockchain information security technology based on B+ search tree algorithm (Item No.: 2020ZDZX3103)

2. “Innovation Research on Cross-border E-commerce Shopping Guide Platform Based on Big Data and AI Technology”, Funded by Ministry of Education Humanities and Social Sciences Research and Planning Fund (18YJAZH042).

## References

1. Zhang, B., Tan, R., Lin, C.J.: Forecasting of e-commerce transaction volume using a hybrid of extreme learning machine and improved moth-flame optimization algorithm. *Appl. Intell.* **12**(05), 1–14 (2020)
2. Liu, S., Liu, X., Yuan, J., et al.: Multidimensional information encryption and storage: when the input is light. *Research* **2021**(01), 1–17 (2021)
3. Borrego, C., Amadeo, M., Molinaro, A., et al.: Privacy-preserving forwarding using homomorphic encryption for information-centric wireless ad hoc networks. *IEEE Commun. Lett.* **23**(10), 1708–1711 (2019)
4. Li, Z., Yang, X., Shen, K., et al.: Information encryption communication system based on the adversarial networks Foundation. *Neurocomputing* **41**(05), 347–357 (2020)
5. Ma, H., Zhang, Z.: A new private information encryption method in Internet of Things under cloud computing environment. *Wirel. Commun. Mob. Comput.* **2020**(6), 1–9 (2020)
6. Zhuang, W.: Research on data encryption technology based on adversarial neural network. *Comput. Eng. Appl.* **49**(4), 5–15 (2019)
7. Zhan, F., Zhang, S.R.: Research on hybrid encryption DAES algorithm based on cloud computing. *Electron. Des. Eng.* **25**(3), 185–189 (2017)
8. Salameh, J., Al-Tarawneh, M.S.: A Secure exchange technique for secret information and encryption key using hybrid system. *Int. J. Commun. Antenna Propag.* **9**(1), 19–26 (2019)
9. Lai, J., Cai, S.: Design of Sino-Japanese cross border e-commerce platform based on FPGA and data mining. *Microprocess. Microsyst.* **80**(07), 103360–103372 (2020)
10. Wang, S.Y., Wu, X.F., Hu, S.G., et al.: A hybrid encryption algorithm based on Logistic chaotic equation and ECC. *Mod. Inf. Technol.* **2**(2), 103–104 (2018)

11. Liu, Y.C., Wang, J., Qu, Q.F.: Text information hiding technique by carrier of song poetry based on hybrid encryption. *Comput. Technol. Dev.* **28**(01), 138–143 (2018)
12. Chen, Y.W., Liu, Y.L., Ma, L.T., et al.: Research on improvement of AES and ECC algorithm and hybrid encryption for substation. *Foreign Electron. Meas. Technol.* **39**(10), 68–73 (2020)
13. Liu, Q.Y., Chen, L., Liu, L.: A hybrid encrypted file protection system based on hardware fingerprint. *J. Shandong Agric. Eng. Univ.* **36**(01), 53–56 (2019)



# Research on Fuzzy Retrieval Method of Blockchain Information Based on B+Tree Index Structure

Jia-hua Li<sup>(✉)</sup>

School of Information Engineering, Guangzhou Vocational and Technical University of Science and Technology, Guangzhou 510550, China  
lijiahua21223@163.com

**Abstract.** Due to the large scale of blockchain information, some traditional retrieval methods are difficult to play a role, so it is necessary to use distributed retrieval method for information retrieval to effectively meet the needs of big data retrieval. Therefore, a blockchain information fuzzy retrieval method based on B+tree index structure is studied. This method first analyzes the problems of blockchain information retrieval, and then uses B+tree index structure to build an index about the target text, and then expands the key words to achieve the fuzziness of the key words and improve the comprehensiveness of the retrieval. Finally, the similarity between the key words and the index is calculated to achieve fuzzy information retrieval. The experimental results show that the retrieval accuracy and comprehensiveness of the research method have been effectively improved, and the research goal has been achieved, and the method can effectively meet the needs of big data retrieval.

**Keywords:** B+tree index structure · Blockchain · Fuzzy retrieval method

## 1 Introduction

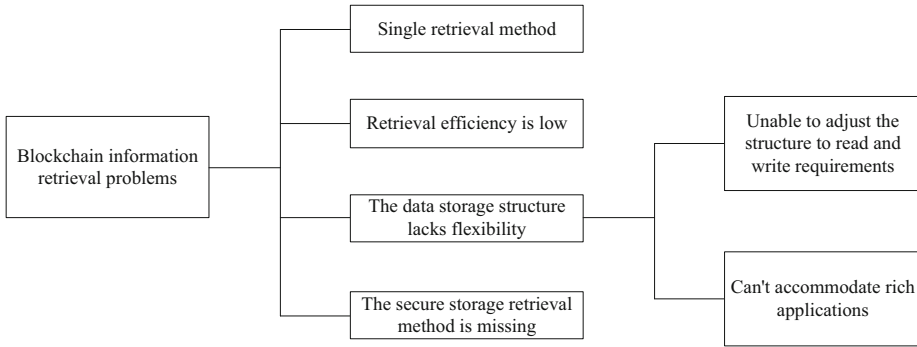
With the rapid development of information technology, more and more data needs to be stored on cloud servers, but the existing cloud storage servers are not completely trusted. In order to realize the confidentiality of data, the data owner encrypts the data and sends the data to the cloud storage server in the form of ciphertext, so that the cloud server cannot obtain the content of the plaintext data, but how to retrieve the information on the cloud server is a difficult problem [1]. Information retrieval is a method by which the user retrieves the required information from the data set. The user generates the query request according to his own needs and sends the query request to the server that has the database, and then the server returns one or more relevant results to the user. With the popularization of the Internet, online information retrieval has become very common in people's life. At the same time, the user's retrieval privacy is more difficult to be protected. Content providers refine their profiles by keeping a

record of what users have searched for. These ISPs can use these user profiles to make money by providing targeted advertising or content recommendations. In some cases, users want their searches to be private. For example, investors looking at the current market value of certain stocks in a stock market database may be reluctant to disclose that they are interested in that stock because this could inadvertently affect the stock price. When a user submits a query request to the server, he hopes that his query information will not be disclosed. Therefore, in order to solve these problems, a variety of new information retrieval methods have been proposed in related fields. For example, some scholars proposed symmetric searchable encryption algorithm for the first time to solve the problem of ciphertext search, but the search efficiency of this scheme is very low based on linear scan. Some scholars put forward the definition of safe index, and use Bloom filter to construct safe index, but there are some errors in the search results. Some scholars have also proposed a searchable scheme of linked keywords, which allows users to retrieve multiple keywords, but the location of keywords needs to be specified in the scheme. In order to further improve the ability of information retrieval, some scholars proposed a single keyword searchable encryption scheme on the blockchain, which realized the data retrieval on the blockchain, but the search results were inaccurate and the search efficiency was low. In order to solve this problem, some scholars proposed a multi-keyword sorting search scheme which supported dynamic updating. In this scheme, the balanced binary tree was used to build the index and the greedy depth-first search algorithm was used to sort the retrieval results.

Based on the predecessors' research experience, based on the current information sharing as the low efficiency of data search and search results are biased, put forward a kind of based on B+tree index structure blockchain fuzzy information retrieval method, in order to use B+tree index structure improve the efficiency of the cipher text search, fully to solve the problems existing in the traditional method, can be widely applied to medical data, and other fields, it has certain contribution to the further development of information technology.

## 2 Problems in Blockchain Information Retrieval

This section focuses on the organization, storage and management of blockchain data, and takes Ethereum as an example to analyze and study the blockchain query mode, from the underlying storage to the abstract data structure, and then to the specific data fields; Then three solutions are summarized and analyzed from different angles; Finally, some ideas for the future development of Block DataBase are put forward, in order to provide some reference for the development of this field. The problems faced by blockchain information retrieval are shown in Fig. 1.



**Fig. 1.** Problems faced by blockchain information retrieval

## 2.1 Single Retrieval Method

The current blockchain system has a single retrieval method, which only supports a limited number of fixed fields as the key matching search, can not support complex queries, is not suitable for rich applications, and can not meet the needs of complex data analysis. Ethereum query specific flow: first, the query request is parsed to get the key; Then we call the Get (key) method of LevelDB to get the corresponding Value. Finally, the RLP deserialization and decoding of the Value are carried out, and then the Json code is encapsulated and returned to the client.

At present, the data field design of blockchain does not consider the query requirements of data analysis at the beginning. It only supports a limited number of fields to meet the basic functions, and does not support user-defined fields. The type is fixed and relatively single, and there is basically no scalability. Moreover, it can not support complex analytical queries such as Range query, Top-k query or fuzzy query. If you want to meet other complex application scenarios or complex query requirements, you need to encapsulate the retrieval layer and build the corresponding index structure in the upper layer of LevelDB, which reduces the attraction of blockchain system to rich applications.

## 2.2 Low Retrieval Efficiency

At present, the underlying storage databases used by mainstream blockchain platforms are mostly key value databases such as leveldb, which are suitable for write intensive business requirements, and the read performance is not excellent. However, the write concurrency of current blockchain systems is not large. For example, the write concurrency of bitcoin transactions is only about 1 transaction per second, and that of Ethereum is about 7 transactions per second. The data storage structure lacks flexibility. When the data in the blockchain system is increasing and the application is expanding, the query demand will increase. At this time, the defects of excessive write performance and insufficient read performance of the underlying storage system become the key factors limiting the query performance.

### 2.3 Data Storage Structure Lacks Flexibility

#### (1) Unable to structure for read and write requirements

The current situation of the Internet is that both the reading and writing needs are taken into account, but the reading and writing of the blockchain system is unbalanced, and it can not meet the flexible adjustment according to the situation. Due to the numerous application fields of blockchain system, their requirements for read-write performance are different. For example, in e-commerce payment system, the transaction throughput will be very large, which requires frequent write operations and relatively few query requests; In the bank's settlement system, the number of transactions per unit time is limited, so the pressure on the write performance will be reduced. However, due to the large number of user queries, the complex audit process and frequent analysis and operation of the salesman, the query is needed, which puts forward higher requirements for the read performance. However, the performance of the existing blockchain system is difficult to meet the requirements of read-write performance adjustment.

#### (2) Unable to adapt to rich applications

At present, the number of fields supported by blockchain data is small, the structure is relatively fixed, and the query processing logic is relatively simple. With the expansion of the scope of application, the data structure of the application will be more complex. The current system can not add and manage user-defined fields, and it is difficult to provide effective support for new applications.

### 2.4 Lack of Secure Storage Retrieval Method

At present, the data of blockchain system is mainly divided into two categories: one is completely open, which can be accessed by anyone and stored in clear text, such as bitcoin, Ethereum and other public chains; One is authorized access, and the data that can be obtained is also stored in clear text. Whether it is internal leakage or hackers' intrusion, the data will be directly exposed.

## 3 Design of Blockchain Information Fuzzy Retrieval Method

With the unique advantages of de trust, weak centralization, data traceability and anti tampering, blockchain is gradually becoming a popular platform for data storage and sharing. Reasonable data organization and management and perfect data privacy protection mechanism are the key factors for the data platform to be recognized. However, due to the decentralized digital currency at the beginning of the design, the data organization form of blockchain is solidified, resulting in single retrieval method and lack of effective data privacy protection design, which hinders the implementation of blockchain in related fields with strong data security requirements and data query requirements, such as smart medicine, financial technology and e-government [2]. The searchable encryption technology can make the data meet the requirements of secure storage and support efficient retrieval, but at present, most of the researches are focused on the searchable encryption in the cloud environment, and there is a lack of searchable Encryption Research for the blockchain data platform. Therefore, this paper studies a

method to improve the efficiency and accuracy of keyword retrieval while achieving the security of ciphertext retrieval. In terms of process, the method starts from the data provider hosting the document data to the cloud server, and then the cloud server returns the query results to the user. There are five steps to complete a cycle of data hosting and query. The specific work flow is as follows:

- (1) The data provider uploads the encrypted data file and the encrypted index corresponding to the data file. The data owner needs to extract the keywords from the file set  $F$ , use the edit distance algorithm to generate the fuzzy keyword set, and construct it as an index file. Then the data provider encrypts the file set  $F$  and stores it in the cloud computing server in the form of ciphertext to protect the data provider's document privacy information. The encrypted document set is  $E$ . In order to retrieve the encrypted document set  $E$ , the data provider needs to establish a retrievable encryption index for  $F$  before encrypting the document set  $F$ . The retrievable encryption index refers to  $I$ . Then the data provider hosts the encrypted document set  $E$  and the corresponding retrievable encryption index  $I$  to the cloud computing server [3].
- (2) When the user queries, the search request  $W_a$  is sent to the data provider through the encrypted connection. In order to query the file data information, the corresponding query encryption index needs to be constructed. Only the data provider can build the retrievable encrypted index and query encrypted index of the document, so the user needs to send the query keywords to the data provider [4]. In order to ensure the security of users' query privacy transmission on the network, data providers and data users communicate directly through encrypted connections.
- (3) When the data provider receives the transmitted search request  $W_a$ , it returns the encrypted query index. The data provider constructs the corresponding query encrypted search request  $T_W$  according to the query keywords sent by the user, and returns it to the data user.
- (4) The user sends the encrypted search request  $T_W$  to the cloud computing server. The user receives the encrypted query request  $T_W$  returned by the data provider and directly sends it to the cloud computing server for query request without further operation.
- (5) The cloud computing server returns the query result. After receiving the user's search request  $T_W$ , the cloud computing server matches the user's search request  $T_W$  with the retrievable encryption index  $I$  under the encryption condition, and then returns the encrypted file name sorted by relevance to the user [5].

The keyword oriented ciphertext retrieval scheme can be roughly divided into three stages: index structure establishment; The key words are fuzzy; Matching retrieval is realized. The following is a detailed analysis of these three processes.

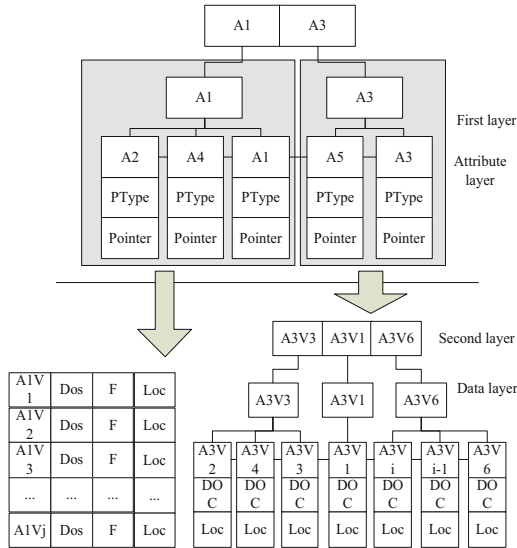


### 3.1 Index Structure Establishment

Index technology is one of the key technologies of modern information retrieval, search application and data mining. Inverted index and B+tree index are two widely used index technologies. Among them, inverted index technology has the advantages of relatively simple implementation, fast query speed and easy to support synonym query. It is widely used in large-scale document set retrieval and other information retrieval. In addition, B+tree is a variant of B tree, which is suitable for the application environment of random and sequential processing, and it has a wide range of applications in database system, file system indexing and so on [6]. In theory, it can index the attributes of any data type, and keep the index level corresponding to the size of the data file. Search has stable I/O overhead, good support for index update, and can bear a variety of workload.

#### Design Idea of Index Structure

The index object contains two kinds of data, namely character data and numerical data. This paper adopts the idea of layering in the design of index structure. That is to say, the first layer of the index structure indexes the attributes to which the data belongs, and the second layer indexes the attribute values corresponding to the data attributes of the first layer. At this time, if the attribute value belongs to character type data, the inverted index is established. If the attribute value belongs to numerical type data, the B+tree index is established. In this way, it is different from the traditional B+tree structure to build B+tree index for all different types of data. The index structure does not establish B+tree index for all character data, so the time cost caused by node splitting can be avoided, and the extra storage space occupied by temporary nodes in the process of node splitting is also relatively reduced. Therefore, the index structure improves the index creation speed and space utilization. Inverted index has good performance for character data retrieval, but it can not meet the requirements of cross data range retrieval for numerical data. B+tree has advantages in numerical data retrieval because of the order of leaf nodes. Therefore, the hybrid index structure in this paper adopts the idea of layered and building different index structures for different types of data to complete the cross data range retrieval, that is, all numerical data are built B+tree index in the second layer. When the cross data range retrieval is needed after the index structure is built, the task will be completed by the B+tree index part of the second layer. The hybrid index structure makes use of and inherits the advantages of the two structures, while discarding the disadvantages of the two structures. In order to achieve efficient index creation speed and space efficiency at the same time, it can complete the requirements of cross data range retrieval. The hybrid index structure is shown in Fig. 2.



**Fig. 2.** Hybrid index structure

In Fig. 2, the first layer is to build a B+tree index structure for the attribute of the index object. In the B+tree index structure of this layer, all non leaf nodes store specific attributes, and all leaf nodes contain three parts of information  $\langle A_i, Ptype, Pointer \rangle$ :

- (1)  $A_i$  is the attribute of the indexed data,  $i \in [1, n]$  and  $n$  are the number of all attributes.
- (2)  $Ptype$  is the pointer type  $Ptype \in \{B+tree, Inver\_index\}$  to the layer 2 structure.
- (3)  $Pointer$  is the pointer to the root node or inverted header of layer 2 B+tree, that is, according to different types of attributes, the pointer points to different index structures.

The second layer is the inverted table or B+tree index structure of the attribute values corresponding to the first layer attributes. The index structure of the inverted table contains four parts of information  $\langle A_iV_j, Doc, F, Loc \rangle$ :

- (1)  $A_iV_j$  is the  $j$  attribute value of the  $i$  attribute,  $i \in [1, n_1]$ ,  $j \in [1, m]$ ,  $n_1$  are the number of character attributes, and  $m$  is the number of attribute values contained in the  $i$  attribute.
- (2)  $Doc$  is the file number of the attribute value, and each file number is unique.
- (3)  $F$  is the frequency of attribute values appearing in the file.
- (4)  $Loc$  is the file location information of the attribute value.

The non leaf nodes of the B+tree index structure in the second layer only contain attribute values, and the leaf nodes are all ordered, and each leaf node contains three parts of information  $\langle A_RV_S, Doc, F, Loc \rangle$ :

- (1)  $A_R V_S$  is the  $S$  attribute value of the  $R$  attribute,  $R \in [1, n_2]$ ,  $S \in [1, p]$ ,  $n_2$  are the number of numerical attributes in the file, and  $p$  is the number of the  $R$  attribute.
- (2)  $Doc$  is the file number of the attribute value, and each file number is unique.
- (3)  $Loc$  is the location information of the file where the attribute value is located.

**Index Structure Creation Process**

The creation process of hybrid index structure is as follows:

- (1) If the file to be indexed is a new attribute of the file, a new node is created for the attribute in the first layer of the hybrid index structure.
- (2) The index structure of the second layer is determined according to the attribute value type of the attribute. If it is a numeric attribute value, a B+tree index is established for the attribute value; If it is a character type property value, an inverted index is established for the property value.
- (3) Repeat step (1). If the current attribute is the same as the previously indexed attribute, no new attribute node is added in the first layer of the hybrid index, and only the attribute value of the attribute is added to the corresponding index structure of the second layer.
- (4) If the current attribute is different from the previously indexed attribute, a new attribute node is added in the first layer of the hybrid index structure, and then according to step (2) to determine what type of index structure needs to be established for the attribute value. Repeat the above steps until all data are indexed.

**3.2 Fuzziness of Key Words Based on Query Expansion**

At present, there are mainly two methods for query expansion, as shown in Table 1.

**Table 1.** Comparison of query expansion methods

Aspect	Middleware method of external database	Built in index optimization method
Query performance	Speed is generally better than the current system	Slightly slower than the current system
Space consumption	At least twice the extra space (full data copy+index structure)	The extra space added can be ignored (index structure of specific fields)
Write performance	Write performance is unchanged because internal and external data is asynchronous	Write performance will decline because it takes time to update the secondary index

(continued)

**Table 1.** (continued)

Aspect	Middleware method of external database	Built in index optimization method
Scalability	It is easy to extend and has good external operability	The internal expansion is relatively complex and needs to modify the blockchain structure, which is not friendly to the old system
Consistency maintenance and fault tolerance	Because of asynchronous operation, consistency is difficult to guarantee; System crash is prone to inconsistency	Compared with native system, consistency maintenance only needs to increase the time of building secondary index

The principles of the above retrieval methods are generally through the accurate matching of keywords to achieve retrieval, that is, the user query keywords need to be completely consistent with the server-side index keywords to retrieve the corresponding documents [7]. In order to improve the flexibility and usability of search, some scholars further proposed the keyword based fuzzy retrieval technology. These fuzzy retrieval schemes are mainly based on the editing distance to measure the similarity of keywords, which can solve the problems of user input errors and inconsistent keyword formats, such as users want to input the keyword “million” but mistakenly input “million”, or input the keyword “data mining” as “data mining”. To sum up, the existing retrieval schemes do not consider some words related to keyword semantics. When the retrieval user does not know much about the domain to be retrieved, the keywords submitted by the user are difficult to express the user’s actual retrieval intention, and some documents containing semantic related words are missed, resulting in incomplete retrieval results and reducing the user’s satisfaction with the retrieval results [8].

For example, if a user wants to search for articles related to football, but he is not familiar with this field, he can only submit the keyword “football” for retrieval, but in the following case, it will cause the problem of missing.

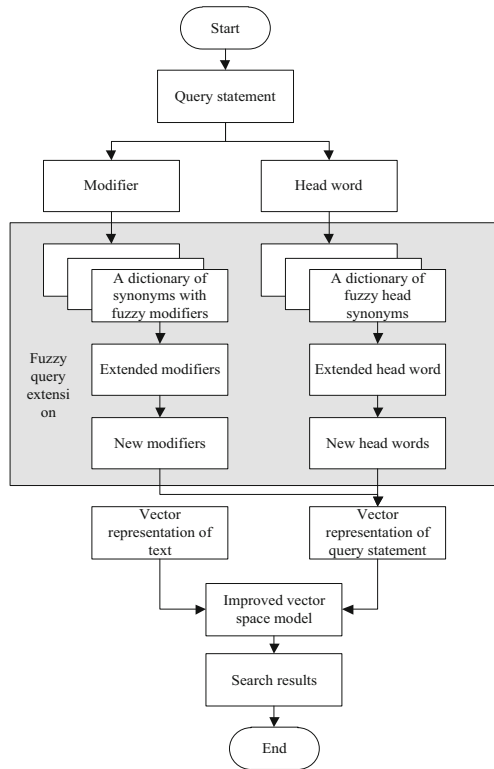
Article 1 is about the content of football transfer market, the key words extracted are “transfer, value, make up”.

Article 2 is about the football game situation, the key words extracted are “score, referee, foul”.

These two articles are related to football, which is also what users want to know. However, the keywords extracted from these documents do not contain the keyword “football”, and these documents will be missed in the retrieval process.

In view of the above shortcomings, this paper realizes the fuzziness of search words based on query expansion, so as to improve the comprehensiveness of retrieval. That is to say, according to the co-occurrence probability of keywords in the document, the semantic relationship database of keywords is constructed. In the retrieval process, the cloud service provider extends the semantic of query keywords submitted by users, Then, the retrieval is performed according to the expanded keyword set, and finally the retrieval results are sorted according to the comprehensive relevance value of the

retrieved documents, and the sorted retrieval results are returned according to the user’s requirements (if the user requires to return the first  $k$ , only the first  $k$  documents are returned) [9]. The basic process of keyword fuzziness based on query expansion is shown in Fig. 3.



**Fig. 3.** Basic process of keyword fuzziness based on query expansion

The specific description is based on the query statement given by the user. Usually it contains the compound phrase with modifier modifier, which is marked by the word character and synonym is used.

In dictionary, we can get the synonyms of modifier and central word respectively. By re combining these extended modifier and central word in statistical sense, we can get new synthetic phrases. The paper constructs vector space model by using the phrase, calculates the correlation between text vector and query vector, sorts the text according to the value of correlation degree, and finally, gets the text related to the semantic of query statement.

### 3.3 Implementation of Fuzzy Retrieval Based on Query Expansion

Boolean retrieval model is the most traditional and mature retrieval model, which has a wide range of applications in the field of information retrieval. Fuzzy retrieval model is an improved product based on Boolean retrieval model and fuzzy set theory. It defines the fuzzy relationship between the query statement and the related literature. The fuzzy retrieval model relates the text information and the query data to a certain extent, and assumes that there is a set of fuzzy text information related to each word in the query statement. In other words, each query word is defined as a fuzzy set, and the elements in the set are the text information for retrieval. Each independent text in the retrieval text information set has a membership degree for each query word in each query statement.

#### Fuzzy Retrieval Based on Word Incidence Matrix

##### (1) Constructing keyword matrix

The word incidence matrix is a word matrix composed of the semantic relationship values between keywords and query words extracted from relevant text information, that is, the elements in the set composed of keywords and query words are taken as rows and columns. Assuming that the word incidence matrix is represented by  $W_{k \times k}$ ,  $k$  represents the number of elements in the set. The element  $w_{ij}$  in the matrix corresponds to the semantic relationship values between words  $i$  and  $j$ , It represents the semantic similarity between two words. In order to make its value range within the interval  $[0, 1]$ , formula (1) is used for calculation:

$$w_{ij} = \begin{cases} \frac{N_{ij}}{N_i + N_j - N_{ij}}, & i \neq j \\ 1, & i = j, \end{cases} \quad (1)$$

Among them,  $N_{ij}$  represents the number of documents containing both  $i$  and  $j$ ,  $N_i$  and  $N_j$  represent the number of documents containing  $i$  and  $j$  respectively. When  $w_{ij}$  is 0, it means that there is almost no semantic correlation between the two words; When  $w_{ij}$  is 1, it means that there is the strongest correlation between the two words.

##### (2) Using keyword matrix to calculate membership degree

For fuzzy retrieval model, each query word corresponds to a set containing fuzzy text information. Let  $D_i$  denote the fuzzy text information set related to word  $V$ , then for any independent text information  $d_j$ , its membership degree  $R_{ij}$  belonging to set  $D_i$  is calculated by formula (2):

$$R_{ij} = \bigoplus_{k \in d_j} W_{ik} \quad (2)$$

Among them,  $W_{ik}$  represents the semantic relation value of word  $i$  and word  $j$ ,  $k$  is the word in independent text  $d_j$ , and  $\bigoplus$  represents fuzzy operator:

$$\bigoplus_i X_i = 1 - \min_i (1 - X_i) \quad (3)$$

Then the membership degree  $R_{ij}$  of any independent text information  $d_j$ , which belongs to set  $D_i$ , can be reduced to:

$$R_{ij} = 1 - \min_{k \in d_j} (1 - W_{ik}) \tag{4}$$

(3) Convert user query

Usually, a query statement containing multiple query words is input to the computer during retrieval. The traditional Boolean retrieval replaces the relationship between these query words with Boolean logic expression. The fuzzy retrieval model uses “truth table method” to transform the query statement into a main disjunctive paradigm composed of minimal items, Let  $Q(h)^-$  denote the set of negative query words and  $Q(h)^+$  denote the set of positive query words in the query statement, then the membership degree  $T_i(h)$  of any independent text information  $d_i$  belonging to the whole query statement is:

$$T_i(h) = 1 - \left( \prod_{j \in Q(h)^+} R_{ij} \right) \left( \prod_{j \in Q(h)^-} R_{ij} \right) = 1 - \left( \prod_{j \in Q(h)^+} R_{ij} \right) \left\{ \prod_{j \in Q(h)^-} (1 - R_{ij}) \right\} \tag{5}$$

where  $R_{ij}$  is the membership of independent text information  $d_i$  to the corresponding query word:

**Word Vector**

Word vector technology is a kind of word representation method through distributed expression of words in the language model. In the training process, words are transformed into dense vectors, and the distance between the corresponding vectors is relatively close for similar words. Therefore, word vector can be used to calculate the similarity between words. Cbow model is a training word vector model proposed by mikolov in 2013. The network structure of cbow model consists of three layers: input layer, projection layer and output layer. The input layer is one hot coding corresponding to the input context. The projection layer sums up the input initial vectors, and the output layer corresponds to a tree structure, which is a tree structure constructed with the words above and below as leaf nodes, and the number of times each word appears in the corpus information as weight. The core idea of cbow model is to predict the word  $w_t$  when the context  $w_{t-2}, w_{t-1}, w_{t+1}$  and  $w_{t+2}$  of the current word  $w_t$  are known.

**Query Extension**

As we all know, in natural language, a word may express several meanings, and the same several different words may express the same meaning. When searching, the computer may not return documents with the same meaning as the query words but different words. Query expansion is to improve query efficiency by adding words with similar semantics to query words [10]. In this paper, we use the word vector trained by CBOW model to calculate the similar words of the query term and expand the query term.

(1) Word segmentation and part of speech tagging are carried out for the input query sentences. In order to avoid the interference of useless information, the stop words are removed from the sentence. The adjectives, adverbs and the query core words modified by them are retained to form the keyword set of the query item.

(2) The word vector of Related words in the keyword set is extracted. The cosine distance of space vector is used to calculate the similarity between words.

(3) Take the first  $N$  words with the largest similarity with the corresponding words as the extension.

Based on the above research, this chapter studies the implementation of fuzzy retrieval based on query expansion, that is, the retrieval is completed by calculating the corresponding similarity between keywords and index words [11, 12]. The specific process is as follows:

Step 1: build a large-scale unlabeled data set, and preprocess it by word segmentation, part of speech tagging and filtering stop words. On this basis, train the processed data set with word vector model (CBOW model) to get the corresponding word vector;

Step 2: construct the text information data set as the retrieval content, and extract 20 keywords from each document (these keywords can generally summarize the content of the article) after the preprocessing operations of word segmentation [13], part of speech tagging and filtering stop words;

Step 3: preprocess the query statement, calculate the similarity according to the cosine distance, expand each word in the query item by  $N$  similar words, and transform the query statement into the main disjunctive normal form of the minimal item;

Step 4: use the extracted keywords and the expanded query words to form the word incidence matrix, get the relationship value between words, go through the whole data set, and calculate the membership of each article belonging to the query sentence;

Step 5: output the corresponding search results after sorting.

## 4 Simulation Experiment

### 4.1 Experimental Environment

This paper uses Java and go language; The experimental results are stored in the file, and the data is drawn into a chart by MATLAB. All the experimental procedures are implemented under the operating system of Ubuntu 16.04 (64bit). The processor platform is Intel (Xeon es-2630 v4.2 GHz \* 40), and the memory size is 250 GB. However, according to the characteristics of memory configuration of Java virtual machine and the system overhead, some key parameters of Java virtual machine are configured as follows: the minimum memory size of `-Xms2048m` is set to 2G, and the maximum memory size of `-Xmx64000` is set to 64G.

### 4.2 Experimental Data

In order to ensure the reliability of the experiment, the 3.2G Wikipedia Chinese corpus, 17901 articles and news corpora including politics, economy, culture, medicine, history and other aspects crawled on the Internet are used as the corpus training set of CBOW



model. 150 articles, totally 1350 articles, are selected from 9 fields of politics, military affairs, economy, culture, medicine, history, sports and so on, retrieve as a text data set.

### 4.3 Evaluation Criterion

Generally, precision and recall are used to evaluate the performance of information retrieval system

- (1) Precision ratio refers to the percentage of the number of relevant documents in the total number of documents, which reflects the retrieval accuracy, and its complement is the false detection rate;

$$\text{Precision ratio : } q = \frac{\alpha}{\beta} \quad (6)$$

where  $q$  is the precision ratio;  $\alpha$  is the amount of relevant information retrieved;  $\beta$  is the total amount of information retrieved.

- (2) Recall rate refers to the percentage of the number of relevant documents detected in the system, which reflects the comprehensiveness of retrieval, and its complement is the missing rate;

$$\text{Recall rate : } p = \frac{\chi}{\gamma} \quad (7)$$

Where  $p$  is the recall rate;  $\chi$  is the amount of relevant information retrieved;  $\gamma$  is the total amount of relevant information in the system.

It can be seen from formula (6) and formula (7) that recall and precision are mutually exclusive. When the recall rate is increased, the precision rate will decrease accordingly; When the recall is reduced, the precision will increase accordingly. In order to judge the retrieval efficiency more accurately, the harmonic average of recall and precision is added here.

$$H = \frac{2 \times q \times p}{q + p} \quad (8)$$

In order to get the experimental data and calculate the relevant proportion, 20 experimenters were asked to select the search results respectively, and the relevant documents of each query were obtained by synthesizing the opinions of 20 people.

### 4.4 Word Vector Training

In this paper, the CBOW model is used to train. The dimension of the word vector is set to 100, the size of the context window is set to 5, the words whose frequency is greater than 1 are retained, and 0.025 is taken as the learning rate of the model. The corresponding word vector is obtained by 100 times of iterative training. Table 2 is the sample of the word vector training results.

**Table 2.** Sampling of word vector training results

Input	Similar words	Vector cosine distance
Happiness	Happy	0.794 743 53
	Happy	0.777 624 21
	Fine	0.761 178 97
	Unforgettable	0.728 432 14
	Pleasure	0.698 920 33
China	Our country	0.852 053 22
	U.S.A	0.774 880 76
	The republic of korea	0.749 371 11
	Britain	0.734 025 57
	Japan	0.728 140 57
Sugar urine sugar	Blood sugar	0.691 834 22
	Pancreatin	0.659 388 69
	Hyperlipidemia	0.598 150 42
	Heart disease	0.532 443 88
	Hypertension	0_512 551 30

It can be seen from Table 2 that by calculating the cosine distance between word vectors, words with similar semantic environment can be found more accurately.

#### 4.5 Comprehensive Comparative Analysis

In the data set of 1350 articles in 9 different fields, the recall rate, precision rate and their harmonic average value are taken as the evaluation indexes, and the threshold value is taken as 1. Fuzzy retrieval based on word association matrix and word vector query expansion are used respectively, and the number of expansion items is added to carry out the retrieval experiment. The experimental results are shown in Table 3.

**Table 3.** Comparison results of comprehensive experiments

Project	Word incidence matrix	Word vector extension
$q$	0.2168	0.2817
$p$	0.5971	0.6122
Harmonic average	0.3181	0.3859

It can be seen from Table 3 that the evaluation result of the improved fuzzy retrieval model by query expansion is better than that of the fuzzy retrieval model based on word incidence matrix, and the recall, precision and the harmonic average of the two are increased by 1.51%, 6.49% and 6.78% respectively, indicating that the query expansion based on word vector can improve the query effect and overcome the problem of “word mismatch” to a certain extent.

### 4.6 Analysis of Query Expansion Number N

In this paper, the key step is to use the word vector to calculate the similar words of the query term, and use the similar words to expand the query term. How to get the most appropriate value of the query expansion number  $n$  to ensure the optimal retrieval efficiency is  $N = 0, 2, 4, 5, 6, 8, 10, 12, 14, 15, 16, 18, 20, 22, 24, 26, 28, 30$  problem worthy of discussion, the experimental results are shown in Table 4.

**Table 4.** Analysis of expansion number N

Number of extensions N	Harmonic average
0	0.32
2	0.34
3	0.37
6	0.40
7	0.41
10	0.42
12	0.42
14	0.44
16	0.45
18	0.46
20	0.47
22	0.46
24	0.46
26	0.45
28	0.45
30	0.45

It can be seen from Table 4 that with the change of parameter  $N$ , the harmonic average of recall and precision also changes  $N \in [0, 15]$ ; When the query expansion number  $N \in [15, 20]$ , the relative harmonic average value gradually reaches the maximum value; When  $N > 20$ , the harmonic average is relatively lower, but basically remains stable. This is because with the increasing of  $N$ , the number of extensions is increasing, but the similarity between the extensions and the corresponding query items is beginning to decrease. Therefore, it is better to take  $N \in [15, 20]$  as the number of the corresponding extensions.

## 5 Concluding Remarks

Because blockchain is designed for decentralized digital currency at the beginning, its data organization form is solidified, resulting in a single retrieval method and lack of effective data privacy protection design, which impedes the implementation of blockchain in relevant fields with strong data security requirements and data query

requirements. For this reason, this paper studies the safe storage and safe and efficient retrieval of sensitive data on blockchain, and proposes a fuzzy retrieval method of blockchain information based on B+tree index structure. The experiment shows that, compared with the traditional retrieval model, the retrieval model studied in this paper improves the fuzzy retrieval effect, can fully solve the problems existing in the traditional methods, promote the development of multiple fields represented by medical big data, and has a certain contribution to the further development of the information technology field. This experiment is a preliminary exploration of this method, and many links need to be further improved. In the future, we will try to take different thresholds for text retrieval on the basis of this paper, expand the training corpus, improve the extraction method of keywords, and improve the retrieval efficiency.

**Fund Projects.** 1. Key scientific research platforms and projects of ordinary universities in Guangdong Province in 2020-Research on cross-border E-commerce blockchain information security technology based on B+search tree algorithm (Item No.:2020ZDZX3103)

2. “Innovation Research on Cross-border E-commerce Shopping Guide Platform Based on Big Data and AI Technology”, Funded by Ministry of Education Humanities and Social Sciences Research and Planning Fund (18YJAZH042).

## References

1. Vale, T., Almeida, E.D.: Experimenting with information retrieval methods in the recovery of feature-code SPL traces. *Empir. Softw. Eng.* **24**(3), 1328–1368 (2019)
2. Marcos-Pablos, S., García-Peñalvo, F.: Information retrieval methodology for aiding scientific database search. *Soft Comput.* **24**(8), 5551–5560 (2018). <https://doi.org/10.1007/s00500-018-3568-0>
3. Kaur, P., Pannu, H.S., Malhi, A.K.: Comparative analysis on cross-modal information retrieval: a review. *Comput. Sci. Rev.* **39**(2), 100336 (2021)
4. Alexandrescu, A.: Optimization and security in information retrieval, extraction, processing, and presentation on a cloud platform. *Inf. (Switzerland)* **10**(6), 200 (2019)
5. Vaz, G.J., Barbedo, J.: An information retrieval system based on multiple portlets: communication between its components. *Int. J. Web Portals* **13**(1), 74–86 (2021)
6. Huang, T., Weng, Z., Liu, G., et al.: HD-Tree: an efficient high-dimensional virtual index structure using a half decomposition strategy. *Algorithms* **13**(12), 338 (2020)
7. Banerjee, P.S., Chakraborty, B., Tripathi, D., Gupta, H., Kumar, S.S.: A information retrieval based on question and answering and NER for unstructured information without using SQL. *Wirel. Pers. Commun.* **108**(3), 1909–1931 (2019). <https://doi.org/10.1007/s11277-019-06501-z>
8. Yang, Z., Yu, H., Tang, J., et al.: Toward keyword extraction in constrained information retrieval in vehicle social network. *IEEE Trans. Veh. Technol.* **68**(5), 4285–4294 (2019)
9. Surya, S., Sumitra, P.: An innovative information retrieval model implementing particle swarm optimization technique. *J. Comput. Theor. Nanosci.* **17**(12), 5613–5617 (2020)
10. Parali, U., Zontul, M., Ertugrul, D.C.: Information retrieval using the reduced row Echelon form of a term-document matrix. *J. Internet Technol.* **20**(4), 1037–1046 (2019)
11. Huang, M.X., Lu, S.D., Xu, H.: Cross-language information retrieval based on weighted association patterns and rule consequent expansion. *Data Anal. Knowl. Discov.* **3**(9), 77–87 (2019)

12. Xie, X.Z., Wang, Z.K., Li, Y.X., et al.: Cross-domain semantic information retrieval based on convolutional neural network. *Comput. Appl. Softw.* **35**(8), 73–78 (2018)
13. Shen, X.T., Lv, L.H.: Research on information retrieval location based on cloud computing. *Comput. Knowl. Technol.* **14**(08), 210–211 (2018)



# Reliability Analysis Method of Multi Area Fault Diagnosis and Location in Power Grid with Missing Information

Jun-ci Tang<sup>1,2(✉)</sup>, Tie Li<sup>1</sup>, Jun-bo Pi<sup>2</sup>, Miao Wang<sup>1</sup>, and Feng Jiang<sup>1</sup>

<sup>1</sup> State Grid Liaoning Electric Power Supply Co., Ltd., Shenyang 110000, China  
tangjunci66888@163.com

<sup>2</sup> State Grid Corporation of China, Beijing 100031, China

**Abstract.** Aiming at the problem of poor positioning accuracy of traditional power grid fault diagnosis methods, this paper studies multi area fault diagnosis and positioning reliability analysis of power grid under lack of information. By analyzing the topology of the distribution network, the location of the fault points in the distribution network is quickly determined, and the fault treatment scheme is given; Based on the lack of information, the simplified structure and mathematical model of key equipment are introduced. The matrix is used to describe the switch state and topological relationship, and the fault diagnosis module based on information loss is designed by using the grid fault diagnosis algorithm based on matrix operation. It realizes the fast detection of multi area fault diagnosis and location, improves the reliability of the detection results, and provides convenience for fast fault location and processing.

**Keywords:** Lack of information · Power grid failure · Fault diagnosis · Fault location

## 1 Introduction

With the rapid development of power system, more and more distributed generation access to the distribution system, so the structure of distribution network is more and more complex. When the distribution network fails, accurately and efficiently finding out the fault point or area is the prerequisite for rapid isolation and power supply recovery as soon as possible, which is of great significance to improve the reliability of the distribution system [1]. In the power grid fault diagnosis, the lack of information technology is used to establish the relevant spatial topology and mathematical calculation model, which can make the fault diagnosis quickly and accurately locate the specific switchgear.

At present, the common fault location of power grid is mainly based on dispatching operation model. This model lacks necessary line simulation, does not provide line load transfer scheme, and the control of the line is completely based on the influence of natural on/off. Because the switch control status and the line relationship are not quantified, the problem solving lacks the relevant mathematical model support, The algorithm is inefficient when the structure of the line changes, and there is no

information about the hierarchical structure of the overhead line, and the description of the power grid structure is not intuitive. Based on the information missing technology, the multi area fault diagnosis and location reliability analysis method of power grid is optimized, and a series of processes such as fault diagnosis model are used to determine the fault point information, and a series of processes such as fault handling operation ticket are automatically generated to ensure the accuracy of fault diagnosis and location results.

For this reason, the relevant scholars have carried on the research and made some progress. In reference [2], a power grid fault diagnosis method based on PMU data and convolutional neural network is proposed. The characteristic gas value is transformed from decimal to corresponding binary, which is represented by two-dimensional data, and the two-dimensional data is used as the input of convolutional neural network to train the optimization model. The fault diagnosis accuracy of this model is high, but the accuracy of fault diagnosis is poor. In reference [3], a power grid fault diagnosis method based on superimposed sparse de-noising automatic encoder and Gru network is proposed. The superimposed sparse de-noising automatic encoder is used to reduce the dimension of the sequence, and the sparse feature expression of the data is obtained. The fault type is obtained by using the time-varying feature of the data extracted by Gru. This method can effectively extract high-dimensional data features and reduce the data dimension, but the accuracy of fault diagnosis is poor.

Aiming at the problems of the above methods, this paper puts forward the reliability analysis method of multi area fault diagnosis and location in power grid under the lack of information, uses the fault diagnosis model to determine the fault point information, and automatically generates a series of processes, such as fault processing operation order, to ensure the accuracy of fault diagnosis and location results.

## **2 Multi Area Fault Diagnosis and Location Reliability of Power Grid**

### **2.1 Multi Area Fault Location Algorithm in Power Grid**

Power grid fault start algorithm is used to detect the time of fault occurrence, and is a prerequisite for fault analysis, diagnosis and processing. The whole power grid fault diagnosis system in the application of the actual power grid, is unattended, real-time operation. The core of the fault start algorithm is the start criterion. When the dynamic data of the system meets the start criterion, the start signal is sent to the fault diagnosis system. An ideal starting criterion is that it can start quickly and reliably in case of system failure and has high starting sensitivity [2]. In normal operation state of power grid, including overload, frequency fluctuation and system oscillation, it will not start reliably. In the digital protection device, according to the different protection principles, the classical electric quantity startup criteria are divided into steady quantity startup criteria and sudden variable startup criteria, and the steady quantity startup criteria are mainly divided into excessive startup criteria and insufficient startup criteria [3]. The

starting criterion of sudden variables is based on the calculation and comparison of the changes between the measured values before and after the fault. Generally, it has higher sensitivity than the starting criterion of steady-state variables. Therefore, the sudden variables are widely used as the starting conditions in the fault diagnosis system. According to the different types of electrical quantity, it can be divided into voltage sudden change starting criterion and current sudden change starting criterion [4]. The four sampling value method is used to extract the current sudden change to realize the judgment of fault starting conditions, as shown in formula (1)

$$\Delta i(n) = \|i(n) - i(n - N)\| - \|i(n - N) - i(n - 2N)\| > I_x \tag{1}$$

where  $i(n)$  is the instantaneous value of phase current sampling at  $n$  time (which can be measured by synchronous phasor measuring device).  $N$  is the number of sampling points in each fundamental frequency period of PMU phase current data. In the calculation of each  $i(n)$ , four sampling values are used, covering the sampling span of two periods, so as to reduce the influence of frequency deviation and system oscillation on the current value to a certain extent.  $I$  is the threshold value of the starting criterion of the current sudden change, and the threshold value is also set to effectively avoid the normal fluctuation of the characteristic quantity in the actual power grid [5]. Because the location problem of fault section in distribution network belongs to 0–1 integer programming problem, the binary information missing optimization algorithm is adopted, that is, the binary number is used to express whether the fault section occurs or not, 0 represents no fault, and “1” represents fault [6]. For BPSO, the update formulas of velocity vector and position vector are as follows:

$$v_{im}^{T+1} = \omega v_{im}^T + a_1 r_1^T (P_{best,im} - x_{iu}^T) + a_2 r_2^T (G_{bestim} - x_{im}^T) \tag{2}$$

where  $v_{im}^{T+1}, x_{im}^{T+1}$  is the velocity and position of the  $m$  bit of the  $T + 1$  generation of information  $i$ ,  $\omega$  is the inertia factor,  $a_1, a_2$  is the acceleration factor, and  $P_{best}, G_{best}$  is the optimal value of the information itself and the group over the ages:

$$\text{Sigmoid}(v_{im}^{T+1}) = \begin{cases} 0.98 & v_{im}^{T+1} > V_{\max} \\ \frac{1}{1 + e^{-v}} & -V_{\max} < v_{im}^{T+1} < V_{\max} \\ -0.98 & v_{im}^{T+1} < -V_{\max} \end{cases} \tag{3}$$

In the process of optimization with information missing algorithm, the information keeps approaching the optimal value. When the convergence condition is satisfied, the more likely the fault section hypothesis is, the more information concentration, that is, the more times the information appears [7]. Therefore, it is proposed that the probability of each fault section set can be represented by the proportion of the occurrence times of each missing information, and the corresponding basic probability distribution function can be formed.



After a fault occurs in the distribution network, the distribution terminal installed at each section switch and tie switch can detect the fault current and upload the fault information to the master station. However, due to the bad working environment, communication impact and other reasons, information distortion or incomplete information often occurs in actual operation [8]. Based on this, a fitness function can be constructed to reflect the difference between the actual received fault information and the expected fault information, and then the BPSO algorithm is used to optimize the solution, and the multiple information corresponding to the minimum fitness function in the missing information and the times of its occurrence in the iterative process are recorded. The fitness function is used as the objective function for optimization:

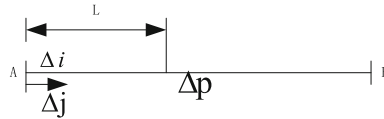
$$F_1(R) = \text{Sigmoid} (v_{im}^{T+1}) \sum_{j=1}^{N_2} |I_j - I_j^*(\mathbf{R})| + \omega \sum_{i=1}^{N_1} |\mathbf{R}(i)| \tag{4}$$

$$I_j^*(R) = \prod D_i - F_1(R) \tag{5}$$

The first traveling wave head transmitted from the received fault point is taken as the initial time to calculate the strongest positive jump point of voltage distribution along the line, and the fault point is determined by the location of the jump point. The calculation formula is as follows:

$$S_i = I_j^*(R) \iiint L - S_w \tag{6}$$

In the formula,  $S_i$  represents the distance between the fault point and the traveling wave bus;  $L$  represents the full length of DC line;  $S_w$  indicates the distance from the jump point to the test point. In order to calibrate the position of the fault point, the first derivative of voltage on the path can be calculated. The maximum positive value of the derivative corresponds to the position of the fault point. To calculate the voltage distribution on the path, the starting time should be calibrated, that is, the time when the initial traveling wave reaches the test point. Traveling wave fault location usually defines the time when the sampling value of voltage traveling wave exceeds a certain value as the initial time. The sampling frequency directly affects the accuracy of initial time calibration. When the traveling wave speed is the speed of light and the sampling frequency is 350 kHz, the maximum fault location error is within 300 m [9]. According to the above fault location principle, the voltage distribution along the line is calculated, and the corrosion fault point of HVDC transmission line is automatically diagnosed. The voltage distribution is the first voltage emission wave at a-terminal and the calculated voltage distribution on the path in the fault component network. Based on this, the principle of grid fault area screening is shown as follows (Fig. 1):



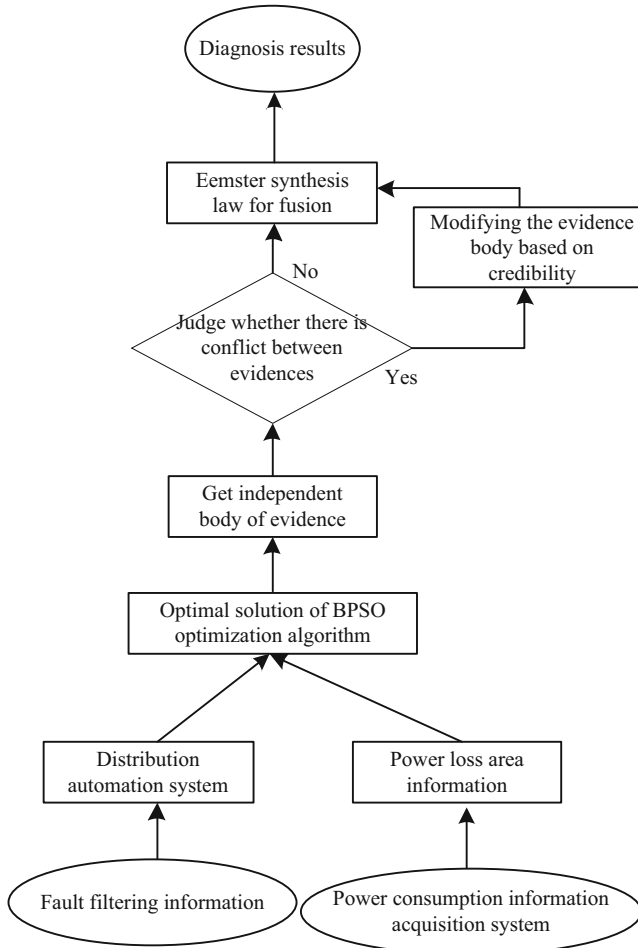
**Fig. 1.** Principle of power grid fault area screening

It can be seen from the figure that  $\Delta i$  and  $\Delta j$  represent the voltage component and current component of the corroded line fault point respectively;  $\Delta p$  is the voltage component of the fault point at  $L$  distance from  $A$  terminal. Under this condition, the voltage component  $\Delta p$  of fault point at  $t$  is calculated as:

$$\Delta p = \frac{1}{2S_i} \left[ \Delta i \left( t - \frac{L}{V} \right) + \Delta j \left( t - \frac{L}{V} \right) C \right] e^{-\frac{rt}{c}} + \frac{1}{2} \left[ \Delta j \left( t + \frac{L}{V} \right) - \Delta i \left( t + \frac{L}{V} \right) C \right] S_i e^{\frac{rt}{c}} \tag{7}$$

In the formula:  $r$  is the resistance per unit length of transmission line;  $C$  is the traveling wave impedance of transmission line;  $V$  is the traveling wave velocity. According to the law of refraction and reflection of traveling wave, the fault point can be diagnosed automatically. When a fault occurs at a certain position of the line, the voltage distribution on the path needs to be calculated. If the voltage component of the section is greater than the total voltage value, it means that the part of the line is faulty, otherwise it is normal [10]. Although the calculated voltage distribution on the path is false at the fault point, the voltage traveling wave is fully refracted at the bus, and it is refracted at the fault point, so the transmitted wave at the opposite end will not change greatly in the propagation process, that is, there is the strongest positive jump at the fault point, and the distance between the jump point and the opposite end is the fault point distance, Fault location can be achieved by using the distribution characteristics of traveling wave voltage [11, 12, 13].

When the fault occurs in different areas, different alarm information will be uploaded. The recognition framework and its basic probability distribution can be obtained from the alarm information. Then the fault section set which can best explain the uploaded information can be obtained by fusing multi-source information with evidence theory, which is the final diagnosis result. In summary, the flow chart of the fault section location method based on multi-source information fusion proposed in this paper is shown in the Fig. 2.



**Fig. 2.** Distribution network fault location process

The proposed algorithm obtains fault over-current information of distribution automation terminal from distribution automation system and abnormal voltage information from power consumption information acquisition system.

## 2.2 Fault Diagnosis of Distribution Network

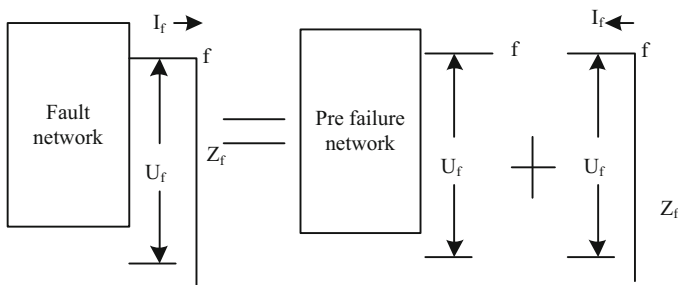
There are many algorithms about fault type discrimination and fault phase selection, but most of them inevitably use sequence component to diagnose fault type more accurately. For example: according to whether there is a zero sequence component, it can distinguish whether it is grounded short circuit or ungrounded short circuit. According to whether there is a negative sequence component in the measured current, it is necessary to distinguish between two-phase short circuit fault and three-phase

fault. It can be seen that the sequence component obtained by symmetrical component method has obvious advantages and importance in power grid fault analysis and diagnosis. Corresponding to the single-phase short-circuit grounding fault, it has the characteristics of positive sequence current, negative sequence current and zero sequence current, which can be represented by the composite sequence network diagram of positive sequence, negative sequence and zero sequence series. Similarly, for other types of faults, different boundary conditions in the form of sequence components will also be obtained. The main characteristics of distribution network missing information fault categories can be summarized as follows (Table 1):

**Table 1.** Main characteristics of distribution network missing information fault category

Type of short circuit	Main features	Composite order network
Three phase short circuit	Only positive sequence components, etc	/
Two phase short circuit	The positive sequence current and negative sequence current of non fault phase are equal	Parallel connection of positive and negative sequence
Two phase short circuit grounding	The non fault phase current is 0	Positive, negative and zero sequence parallel connection
Single phase short circuit	The three sequence currents of the fault phase are equal	Positive, negative and zero sequence series connection

The fault location based on voltage sag information is based on the different position fault, the short-time voltage sag value produced at each node of the system is different. Based on this, it is assumed that each node has a fault in turn, and the fault current at the fault point can be obtained by using the voltage sag value and system impedance. If the node is the real fault point, the current should be consistent with or almost the same with the actual fault current value. Therefore, it can be judged whether each node is a real fault node. After the distribution network fails, it is equivalent to adding an injection current meter at the fault point to divide the distribution network in the fault into two parts, as shown in Fig. 3:



**Fig. 3.** Fault component missing information diagnosis

When the distribution automation terminal at the key position is missed or misreported, the fault location method of traditional single information source will often lead to misjudgment. For example, when the distribution terminal in the fault section fails to report or the adjacent downstream distribution terminal misalarm, it will get the wrong conclusion whether using matrix algorithm or artificial intelligence algorithm. This is the inherent defect of fault section location by using fault overcurrent information alone, and the multi-source information fusion can solve this problem well. It is assumed that there is a fault on Sect. 11, but FTU on Section II fails to upload the flow information to the control center of the main station due to communication reasons, that is, the fault over-current information field is 100100010000000, and the load point information field called by ceias is 0000010. In this case, the fault section set and its basic probability distribution obtained by BPSO algorithm are shown in the Table 2.

**Table 2.** Section diagnosis results in case of loss of fault information of key position

possibility	Fault section set	M1	M2	Fusion results	K value
1	8	0.6772	0.0000	0.5032	0.1276
2	10	0.1381	0.0000	0.0415	
3	11	0.1305	0.9221	0.6946	
4	12	0.0287	0.0000	0.0086	
5	4,8	0.0260	0.0000	0.0079	
6	3,9	0.0000	0.0227	0.0159	
7	4,9	0.0000	0.0191	0.0135	
8	9,11	0.0000	0.0183	0.0128	
9	8,11	0.0000	0.0183	0.0128	

When there are multiple distribution automation terminals that fail to report or misreport, the accuracy of traditional section positioning method is not high, and the optimization algorithm is often used to solve the problem, which often needs iteration many times to obtain the final result. The voltage value at any node in the network can be divided into two parts, the normal component and the fault component are

$$U_i = \Delta p U_i^{01} - Z_i I_f \tag{8}$$

Suppose there is a fault in Sect. 13, but the terminal at the head end of Sects. 8 and 12 does not upload the over-current information to the master station control center due to equipment reasons, that is, the actual uploaded fault over-current information is 100100010, and the load point information field called by CEIAS is 00100. The fault section set and its basic probability allocation M1 and M2 obtained by BPSO algorithm are shown in the Table 3.

**Table 3.** Diagnosis results of multiple distribution automation terminals with information loss

possibility	Fault section set	M1	M2	Fusion results	K value
1	4	0.4586	0.0000	0.1459	0.0751
2	4,9	0.4152	0.0000	0.1342	
3	9	0.0367	0.0000	0.0112	
4	4,12	0.0326	0.0315	0.0348	
5	4,13	0.0315	0.0000	0.0095	
6	13	0.0000	0.8503	0.5953	
7	9,12	0.0000	0.0587	0.0398	
8	12	0.0000	0.0369	0.0259	
9	8,13	0.0000	0.0358	0.189	

The intelligent fault location depends on the fault information fed back by the monitoring center. According to the operation status of the master station before and after the fault, it makes a comprehensive judgment to determine the fault area. After the fault occurs in the distribution network, the distribution automation feeder terminal unit collects the corresponding current waveform and uploads it to the master station for data monitoring. After receiving the feedback information, the switch displacement information on the circuit breaker will control the start of fault location software.

Since genetic algorithm can't directly convert parameters into data, it is necessary to encode parameters into binary digital strings. The genetic operation only operates on the digital string, while the fault location of distribution network should be 0/1 coded according to the requirements of genetic algorithm. The fault of current equipment flowing through the fault line is set to 1, and the fault of current equipment flowing through the fault line is set to 0. By analyzing the status of feeder terminal unit in distribution automation, it is determined whether the switches in each area can be started normally, which reflects the relationship between each element and switch in distribution network. For a single power supply circuit, the switching function can be defined as:

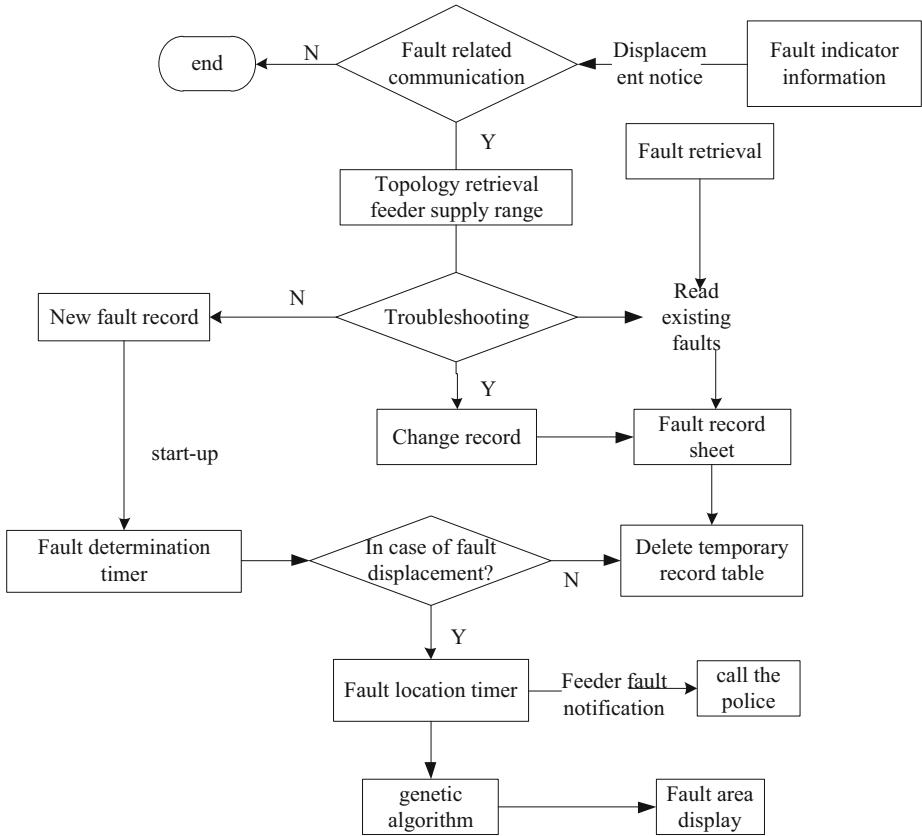
$$F(x)=U_i \left[ 1 - \sum_{m=1}^{m \neq n} x(m) \right] \|x(m+1)\| \tag{9}$$

In the formula:  $F(x)$  is the switch function;  $x(m)$  indicates the state value of the  $m$  element downstream of the switch;  $G$  is the number of elements downstream of the switch. The fitness function is constructed as follows:

$$k(x)=\sum_{j=1}^h |f(x) - F(x)| \tag{10}$$

where  $k(x)$  is the fitness value in space;  $f(x)$  is the real state of the switch. Fault location is calculated by searching the minimum difference between the actual state

value and the derived state value in  $k(x)$  space, which is the minimum solution of the formula. According to the above content, the specific flow of fault intelligent positioning system is shown in the Fig. 4.



**Fig. 4.** Intelligent fault diagnosis process of multi area power grid

The power system of distribution network is a complex system, and there are uncertainties between data and signal. Genetic algorithm can improve the uncertainty of data. Combined with the specific implementation process shown in the figure, the function design of system central station is completed to ensure the effect of multi area fault intelligent diagnosis.

### 2.3 Reliability Optimization of Fault Diagnosis and Location in Power Grid

During the operation of power grid equipment, lidar will scan a set of point data sets in radar coordinate system. The relevant target information points are extracted from the point data set, including scan point, start point and end point. The purpose of feature extraction is to find fault location information points in all information points.

After scanning point data set  $(a_i, \theta_i)$ , most of the feature points are invalid due to reflection, so it needs to be preprocessed. The specific steps of pretreatment are as follows:

Filter point  $a_i > D_T$  data in all point data sets to obtain effective point set  $(a_j, \theta_j)$ ;

Calculate the distance  $a_{j,j+1}$ ,  $a_{j,j+1} = |a_j - a_{j+1}|$  between the two points in the relative concentration of effective feature points;

According to the formula, we can judge whether  $a_j$  and  $a_{j+1}$  belong to the same effective subset  $W_i$ :

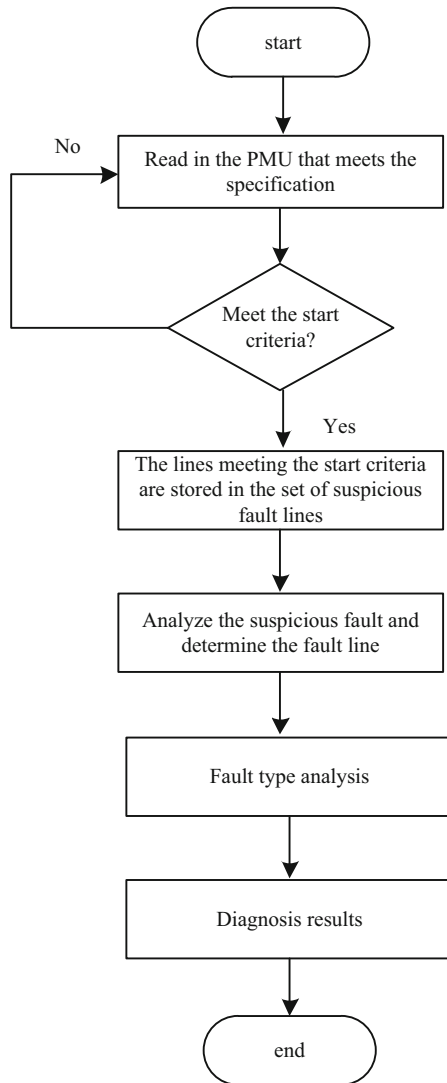
$$a_{j,j+1} \leq A_0 + D_{\min} \cdot \frac{\tan \alpha \sqrt{2(1 - \cos \beta)}}{k(x) \cos \frac{\beta}{2} - \sin \frac{\beta}{2}} \tag{11}$$

Among them,  $D_{\min} = \min\{a_j, a_{j+1}\}$ ;  $\beta$  is the angle resolution of lidar;  $\alpha$  and  $A_0$  are system parameters. Based on this,  $\alpha$  is further calculated as follows:

$$\begin{aligned} v &= \sqrt{a_j^2 + a_{j+1}^2 - 2a_j a_{j+1} \cos \beta} \\ \alpha &= \frac{\pi - \beta}{2} - \arcsin\left(\frac{a_{\min} \sin \beta}{v}\right) \end{aligned} \tag{12}$$

In the formula:  $v$  represents the distance between adjacent feature points, and the measured distance values of these two points are  $a_j$  and  $a_{j+1}$  respectively. If  $a_{j,j+1}$  satisfies the formula, then we need to put  $(a_j, \theta_j)$  into the effective subset  $W_i$  and  $(a_{j+1}, \theta_{j+1})$  into a new effective subset  $W_{i+1}$ . We can judge  $a_{j+1,j+2}$  until the effective point data is completely judged and identified. Based on the power grid fault diagnosis method based on PMU data, a power grid fault diagnosis system based on PMU data is designed and developed by using VC++ 6.0. The diagnosis process is shown in the Fig. 5.





**Fig. 5.** Process optimization of fault location and diagnosis in power grid.

As shown in the figure above, the power grid fault diagnosis method developed in this paper will start from the data, monitor and refresh the received synchronized phasor data in real time, calculate and analyze them, and judge whether they are fault start criteria. If it is satisfied, it is considered that the system may have a fault, and the set of lines meeting the start criteria is stored in the set of suspicious fault lines. Then, the fault location analysis of the lines in the suspicious line fault set is carried out based on their respective data, and the fault lines are quickly selected and stored in the fault line set. The data of the fault line is extracted, and the characteristics of its electric

quantity change are analyzed to diagnose the fault type and fault phase. Finally, a fault diagnosis conclusion including fault occurrence time, fault component, fault type and fault is output. According to the coordinates of the substation equipment in the global coordinate system, the positioning workflow is designed. The specific steps are as follows:

- ① Power on self check;
- ② Send get\_Health request;
- ③ Check whether the data receiving time-out, if the time-out, the broadcast communication error; If there is no timeout, proceed to the next step;
- ④ Check whether protective shutdown is selected. If it is stopped for the first time, it is necessary to send a reat request and return to step 2 with a delay of 5 ms. If the machine is shut down for many times, troubleshooting is needed. If there is no downtime, send scan request;
- ⑤ Receiving ranging data;
- ⑥ Whether the waiting time is out of time. If it is out of time, it is necessary to deal with the fault and check whether the scanning equipment is running normally. If there is no timeout, it is necessary to preprocess the ranging data;
- ⑦ Check if the scan is finished? If it is, it needs to send a stop request to obtain the specific location of the fault. If not, wait until the scan is complete.

According to the positioning workflow, the relevant fault characteristics are extracted and the substation equipment is controlled in real time, so as to realize the fault location and accurate diagnosis of the substation equipment, and improve the reliability of fault location and diagnosis results.

### 3 Analysis of Experimental Results

In order to improve the reliability of this method, the experimental verification analysis is carried out. Taking the ungrounded system of 20 kV power grid in a certain area as the experimental environment, the experiment is built in Matlab environment, in which the hardware environment is Intel Core 3-5501g memory and the operating system is Windows 8. Suppose that a large number of network communication nodes are distributed at 2000 m in the optical fiber environment  $\times$  2000 m uniform array area, written according to the requirements of C++, can add multiple functions of Microsoft, including the system running through the whole windows program, user interface and file operation. The experimental parameters are shown in the Table 4.

**Table 4.** Experimental parameter setting

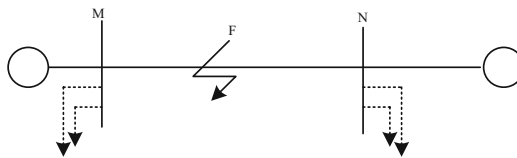
Parameter	Remarks
Optical fiber communication frequency band	3 kHz–8 kHz
Carrier frequency time width	3 ms
Normalized initial frequency of data mining	0.15 Hz
Number of sampling points	256
SNR range	–15 dB–15 dB

According to the experimental environment and parameter setting results, the experimental content is analyzed. Take the first line and the sixth line, i.e. line 1 and line 6, as an example, input line 1 and line 6 into “address input”, i.e. x10001 and x60001. Click the three-phase current button to collect the three-phase current and zero sequence current of the line, as shown in the Table 5.

**Table 5.** Experimental data acquisition

Control area	The first line			Route 6		
Address input	X10001			X60001		
Configuration setting	21			21		
Phase current	A	B	C	A	B	C
	1.2	4.5	3.7	1.1	3.6	5.6
Zero sequence current	3.15 A			3.0A		
Reply message	443210856131349796223416464 1674885620112336544852552555 F5655W15241G12121121QQ111			44321085613134569897414648413 125847994546884555 + 55454545F 5655W15241G12121121QQ111		

It can be seen from the table that the three-phase current and zero sequence current values viewed from the reply information are the original three-phase current data collected from the interface of the master station control center. The simulation model of 500 kV transmission line is built by pscadbmtdc software. The transmission line adopts distributed parameter model, the length of the transmission line is  $L = 200$  km, and the measuring point is located at the M end. The sampling frequency is 1 MHz. The grounding fault is set at the far end and near end of the transmission line Mn respectively, and the fault point is represented by F. In order to simulate the real application environment of the location method, noise is added to the extracted transient traveling wave (Fig. 6).



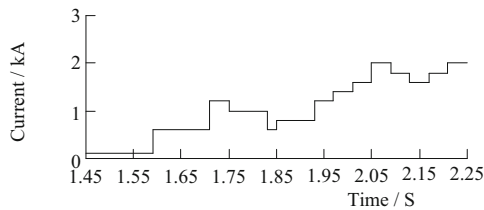
**Fig. 6.** Transmission line simulation model

Based on the above model, the structure parameters of network equipment are further standardized, as shown in the following Table 6.

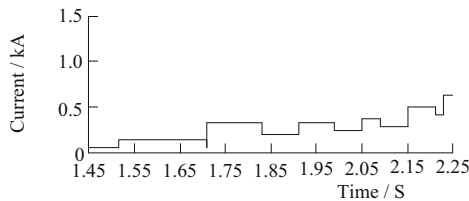
**Table 6.** Structure of power grid equipment

Bus structure	I	II	III	IV
M-end reflection coefficient	<0	<0	>0	>0
N-terminal reflection coefficient	<0	>0	<0	>0
M-terminal bus type	Class I	Class I	Three classes of initials	Three classes of initials
N-terminal bus type	Class I	Three classes of initials	Class I	Three classes of initials

When the bus is a class II bus, there is no reflection wave when the traveling wave reaches the class II bus because the wave impedance is constant. In addition, from the practical application point of view, the class II bus is almost absent, so the case of class II bus is not considered in the simulation example. In order to simplify the analysis process, the influence of traveling wave from adjacent bus can be ignored. The following experiments are carried out to solve the problem diagnosis technology of flexible DC transmission line. The length of the experimental line is 1000 km, and the grounding fault in the positive direction of current occurs at 400 km away from the rectifier side, and the fault time is 2.0S. Based on this, the experimental contents are compared and analyzed (Fig. 7).



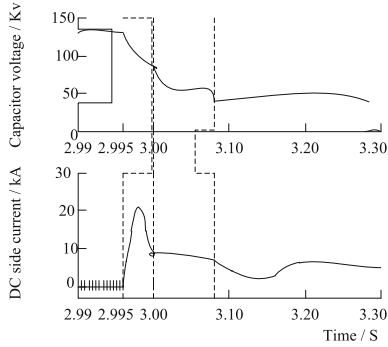
(a) Current before fault



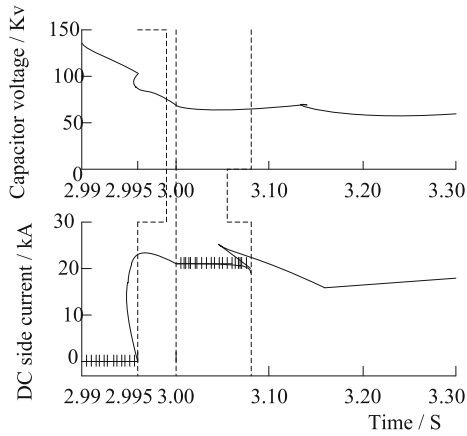
(b) Post fault current

**Fig. 7.** Current diagnosis results before and after fault

According to this point, the rationality of diagnosis technology is compared and analyzed, and the specific comparison content is as follows. The traditional diagnosis technology and automatic location based on traveling wave voltage distribution characteristics are analyzed, and the results are shown in the Figs. 8 and 9.



**Fig. 8.** Fault location results of traditional method



**Fig. 9.** Fault location results of this method

The capacitor voltage obtained by traditional diagnosis technology is also inconsistent with the DC circuit obtained by traditional diagnosis technology based on traveling wave voltage distribution characteristics. Furthermore, the diagnosis errors of traditional diagnosis technology and automatic diagnosis technology based on traveling wave voltage distribution characteristics are compared and analyzed, and the results are shown in Table 7.

**Table 7.** Comparison and analysis of diagnosis error of two technologies

Times	Traditional technology	This paper introduces the technology
1	0.6	0.2
2	0.7	0.1
3	0.8	0.1
4	0.6	0.05
5	0.5	0.1

It can be seen from the table that: under these five experimental conditions, the automatic diagnosis technology based on traveling wave voltage distribution characteristics has relatively small diagnosis error compared with traditional technology, which can better improve the reliability of fault diagnosis and location, and fully meet the research requirements.

## 4 Concluding Remarks

More and more distributed generation connected to distribution system has become the development trend of power grid in the future, but the existence of distributed generation will affect the size and direction of fault current, so the traditional fault diagnosis method of distribution network is no longer applicable. The fault diagnosis method of distribution network based on fault power direction criterion and information missing makes full use of both upload and measured fault information to ensure the accuracy and effectiveness of fault diagnosis of distribution system with distributed generation and improve the speed of fault diagnosis. In the complex situation of information loss or distortion, this method can still accurately determine the fault area of distribution network, which meets the needs of modern distribution network fault diagnosis.

**Acknowledgement.** This paper is supported by the “Research and demonstration application of key technologies of intelligent control robot assistant (SGTYHT/19-JS-215)”, the project of headquarters management science and technology of State Grid Corporation of China.

## References

1. Hassan, Z., Amir, A., Selvaraj, J., et al.: A review on current injection techniques for low-voltage ride-through and grid fault conditions in grid-connected photovoltaic system. *Sol. Energy* **207**(47), 851–873 (2020)
2. Ding, X., Zhu, X., Li, X., et al.: Research on power grid fault diagnosis method based on PMU data and convolutional neural network. In: 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2). IEEE (2020)
3. Ling, Z., Yan, P., Chen, F., et al.: Power grid fault diagnosis method based on stacked sparse denoising auto-encoder and GRU network. *J. Phys. Conf. Ser.* **15**(42), 41–56 (2020)
4. Kruse, J., Schafer, B., Witthaut, D.: Predictability of power grid frequency. *IEEE Access* **8**(1), 149435–149446 (2020)
5. Kortenbruck, J., Premgamone, T., Ortjohann, E., et al.: Asymmetrical grid control with power electronic regulator in distribution level. *IET Renew. Power Gener.* **14**(15), 2830–2839 (2020)
6. Yang, H., Meng, C., Wang, C.: A probability first memetic algorithm for the dynamic multiple-fault diagnosis problem with non-ideal tests. *Memetic Comput.* **12**(2), 1–13 (2020)
7. Gashteroodkhani, O.A., Majidi, M., Etezadi-Amoli, M., et al.: A hybrid SVM-TT transform-based method for fault location in hybrid transmission lines with underground cables. *Electr. Power Syst. Res.* **170**(5), 205–214 (2019)

8. Saffarian, A., Abasi, M.: Fault location in series capacitor compensated three-terminal transmission lines based on the analysis of voltage and current phasor equations and asynchronous data transfer. *Electr. Power Syst. Res.* **187**(3), 106457 (2020)
9. Wu, F., Sun, J., Zhou, D., et al.: Simplified Fourier series based transistor open-circuit fault location method in voltage-source inverter fed induction motor. *IEEE Access* **8**(1), 83953–83964 (2020)
10. Khoramabadi, H.R.S., Keshavarz, A., Dashti, R.: A novel fault location method for compensated transmission line including UPFC using one-ended voltage and FDOST transform. *Int. Trans. Electr. Energy Syst.* **30**(1), 1–31 (2020)
11. Ebrahim, M.A., Aziz, B.A., Nashed, M., et al.: A novel Hybrid-HHOPSO algorithm based optimal compensators of four-layer cascaded control for a new structurally modified AC microgrid. *IEEE Access* **9**(1), 4008–4037 (2021)
12. Tian, S., Li, Z., Li, H., et al.: Active control method for torsional vibration of DFIG drive chain under asymmetric power grid fault. *IEEE Access* **8**(6), 155611–155618 (2020)



# Reliability Analysis of Real Time Operation State of Power System Based on K-Nearest Neighbor

Tie Li<sup>1</sup>(✉), Jun-ci Tang<sup>1</sup>, Jian-ming Yu<sup>2</sup>, Dai Cui<sup>1</sup>, and Di Jiang<sup>1</sup>

<sup>1</sup> State Grid Liaoning Electric Power Supply Co., Ltd., Shenyang 110000, China  
litie9982543@163.com

<sup>2</sup> Beijing Kedong Power Control System Co., Ltd., Beijing 100031, China

**Abstract.** In order to better guarantee the real-time operation state of power system, a k-nearest neighbor based reliability method for real-time operation state of power system is proposed. The k-nearest neighbor technology is used to collect the information of the real-time operation state of the power system, and the evaluation algorithm of the system operation state is optimized according to the collection results. According to the calculation results, the reliability of the real-time operation state of the power system is analyzed, so as to better guarantee the security and stability of the power system operation.

**Keywords:** k nearest neighbor · Power system · Operation reliability

## 1 Introduction

People's daily life and work can not be separated from electricity, in daily life, production need the support of electricity, otherwise it is difficult to carry out normal life and work. Power belongs to energy, and to maintain the balance of power demand, the most important task is to carry out reasonable power dispatching. Using k nearest neighbor technology, we can master and evaluate the faults that people have during the power consumption period, and then carry out the power distribution pertinently, which will bring valuable reference for the establishment of a sound power operation management system. Information management of distribution network is very important, and the composition of the system is also extremely complex. Distribution network should not only be responsible for the operation and management of power enterprises, but also supervise the operation of the power grid, so as to ensure the operation effect of the distribution network [1, 2]. The state model is more practical and simple than the traditional model. It can fully demonstrate the Markov nature of state space transfer of power system. That is, with the support of sufficient sample number, the transfer probability matrix of system state is counted out, and then the current system state and matrix are used, Thus, the change of the probability of system state is analyzed and obtained, which is convenient for power system planning, operation and maintenance. With the advent of big data era, big data has been widely used in power system, and plays a great role in ensuring reliable and stable operation of distribution network.



Therefore, the relevant staff should analyze the reliability of distribution network operation based on big data technology.

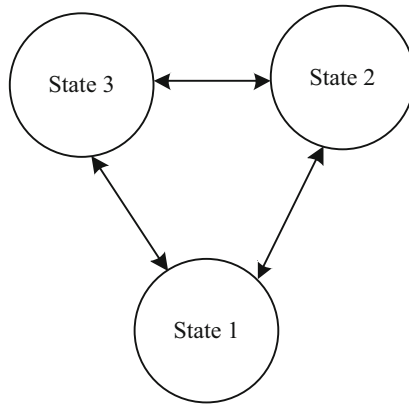
Relevant scholars have made some progress in this field. According to the method of reference [2], an optimization method of data manipulation in power system estimation loop is proposed. The operation state of power system is captured by phase capacity measurement, and the bad data parameters of power system fault are simulated by model compensation method to realize the optimization of data manipulation in power system operation capture. This method can reduce the data noise, But the reliability analysis of power grid data operation state is not accurate. According to the actual demand of multi-source interconnected power system, a four area multi-source interconnected power system including fire, water, gas and renewable energy is established. PID controller based on gray wolf optimization algorithm is used for load frequency control, This method can realize the optimization of load frequency control, but the security of power system operation is poor.

To solve the above problems, this paper proposes a k-nearest neighbor based real-time operation state reliability analysis method of power system, which can effectively improve the security and stability of power system operation.

## 2 Reliability Analysis of Real Time Operation State of Power System

### 2.1 Real Time Operation State Information Collection of Power System

At present, in the process of power system operation reliability evaluation, the electric power department of our country often carries on the related operation according to different types of reliability criteria, and divides different states. Based on this, in order to further promote the effective development of power system operation reliability evaluation, it is necessary to effectively collect the real-time operation state information of power system [3]. Generally, in the actual operation process, operators often divide the operation status of power system into two kinds: normal state and risk state. With the further deepening of related research, the normal state of power system is further divided into healthy state and critical state. Overall, the detailed performance of power system operation state division plays a strong guiding role in security control. Based on this, in the process of related problems analysis, the power system operation state is divided into three kinds: normal state, accident state and risk state [4]. The so-called "normal state" means that there is no component failure in the process of power system operation, and it meets the relevant requirements of "N - 1". The collection and constraint processing of component failure information in the process of power system operation will have a certain impact on the security and stability of the system. Based on this, the state space of power system operation three state model is set to  $E = \{1, 2, 3\}$ . Based on this, the collection system of power system operation state characteristics is optimized as follows (Fig. 1):



**Fig. 1.** Collection system of power system operation state characteristics

Combined with the method of principal component analysis and the reliability data generated in the process of historical operation, the characteristics of various data are selected to reasonably evaluate the indicators of distribution operation stability [5]. Determine the influencing factors, and then regard it as an important part of the artificial neural network, use important standards as the output content, and create a perfect prediction model on the premise of training. The real-time data is transmitted in the model, and then the prediction is made to obtain the standard value of operation reliability in the following specific time period. In addition, the state transition probability matrix is as follows:

$$P = \begin{bmatrix} 1 - p_{12} - p_{13} & p_{12} & p_{13} \\ p_{21} & 1 - p_{21} - p_{23} & p_{23} \\ p_{31} & p_{32} & 1 - p_{31} - p_{32} \end{bmatrix} \tag{1}$$

The power system state transfer algorithm is further obtained as follows:

$$\pi(i) = \begin{bmatrix} \pi_1(i) \\ \pi_2(i) \\ \pi_3(i) \end{bmatrix}^T = \begin{bmatrix} P[X(t_i) = e_1] \\ P[X(t_i) = e_2] \\ P[X(t_i) = e_3] \end{bmatrix}^T \tag{2}$$

In general, row vector  $X(t_i)$  refers to the state distribution of power system at time  $t_i$ . Assuming that the initial state of the power system is  $\pi(i)$ , we can know that based on the definition of power system state transition probability matrix  $P$ , through  $e$  After  $\Delta t$ , the state of the power system changes to:

$$\pi(m) = \pi(m - 1)\Delta tP = \pi(0)P^m \sigma_X \tag{3}$$

Through the understanding and mastering of the current power system state transfer probability matrix  $\pi$ , the operators can realize the understanding and mastery of the steady-state value of the power system state distribution after every  $4t$  time interval in

the future based on the above knowledge and mastery in the actual operation processing process [6]. When the number of time intervals  $n \rightarrow \infty$ , the operation state of the power system will be kept in a certain stable value area, and this value is called the probability of stable state of power system, or the probability of long-term state. The expression of the probability of stable state of power system is as follows:

$$\pi(\infty) = \lim_{n \rightarrow \infty} \pi(n) = \lim_{n \rightarrow \infty} \pi(0)P^n \tag{4}$$

The core connotation of stationary state probability is that the probability of power system in state  $i$  will be close to that in state  $\pi_i(\infty)$  after a long period of operation. Based on the above formula, the stationary state probability of the system can be further derived:

$$\sum_{i=1}^n \pi_i(\infty) = 1 \tag{5}$$

The availability of power system can be further defined as the probability that the power system is in an acceptable state at  $t_i$ :

$$\begin{aligned} A(i) &= \sum_{i=1}^n \pi_i(\infty) \sum_{j \in W} \pi_j(i) \\ &= \pi_1(i) + \pi_2(i) \\ &= 1 - \pi_3(i). \end{aligned} \tag{6}$$

Since the traditional planning reliability evaluation reflects the long-term reliability level of the system, the average value of the long-term statistical data is used for the failure probability and failure rate of components, as shown by the dotted line in the Fig. 2.

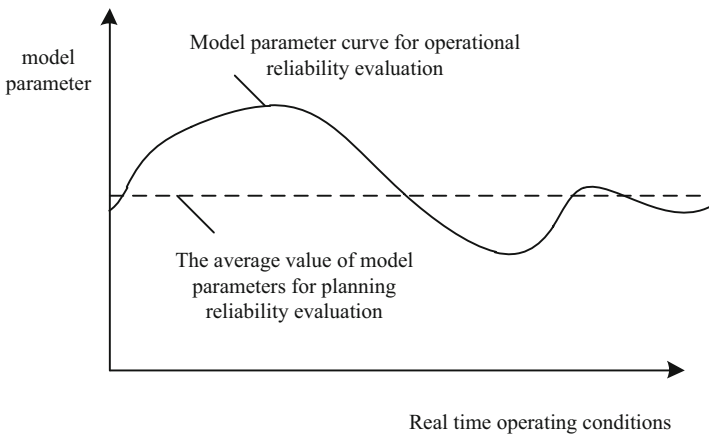
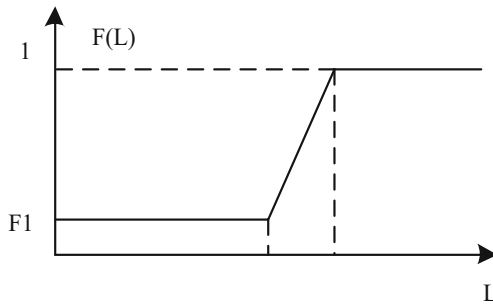


Fig. 2. Characteristic parameter curve of reliability model

In recent years, the power grid accident records show that this kind of fault often occurs, causing the theory seriously divorced from practice. The reason is that there is a problem in the reliability modeling of components, and the outage probability of components should change with the change of system operation conditions [7]. In the actual statistical work of line reliability data, the outage probability of lines is not classified according to the size of power flow, so the relationship between outage probability and power flow cannot be obtained. In this paper, according to the following assumptions, the curve  $F(L)$  of line outage probability changing with power flow is fitted, as shown in Fig. 3:



**Fig. 3.** Curve of power system operation changing with power flow

When the line power flow is within the normal range, the influence of power flow on line outage probability is very small. The line outage probability  $F(L)$  is taken as the statistical value  $F$  of line outage probability, as follows:

$$F(L) = A(i) - \bar{F}_1 \quad (L_{\min}^{\text{normal}} \leq L \leq L_{\max}^{\text{normal}}) \tag{7}$$

where  $L_{\min}^{\text{normal}}$  is the lower limit of the normal value of line power flow and  $L_{\max}^{\text{normal}}$  is the upper limit of the normal value of line power flow. When the line power flow is between the normal value and the limit value, the probability of line fusing or protection device action increases with the increase of line power flow, and the line outage probability is fitted by straight line, as shown in the formula:

$$F(L) = \frac{1 - \bar{F}_1}{L_{\max} - L_{\max}^{\text{normal}}} \times F(L) + \frac{\bar{F}_1 \times L_{\max} - L_{\max}^{\text{normal}}}{L_{\max} - L_{\max}^{\text{normal}}} \sigma_X \tag{8}$$

When the generator frequency is within the normal range, the frequency has little effect on the generator outage probability. The generator outage probability  $F(FG)$  is taken as the statistical value  $\bar{F}_g$  of the generator outage probability, as shown in the following formula.

$$F(F_G) = F(L)/\bar{F}_g - 1 \quad (9)$$

$$\left( F_{G,\min}^{\text{nomal}} \leq F_G \leq F_{G,\max}^{\text{nomal}} \right)$$

where  $F_G$  is the normal value of generator frequency;  $F_{G,\min}^{\text{nomal}}$  is the lower limit of normal value of generator frequency;  $F_{G,\max}^{\text{nomal}}$  is the upper limit of normal value of generator frequency. When the generator frequency exceeds the limit value, the generator protection device acts, and the generator outage probability is 1, as shown in the following formula

$$\begin{cases} F(F_G) = 1 & (F_G \geq F_{G,\max}) \\ F(F_G) = 1 & (F_G \leq F_{G,\min}) \end{cases} \quad (10)$$

where  $F_{G,\max}$  is the upper limit of generator frequency;  $F_{G,\min}$  is the lower limit of generator frequency [8]. When the generator frequency is between the normal value and the limit value, the action probability of the generator protection device increases with the frequency approaching the limit value.

$$F(F_G) = \frac{1 - \bar{F}_g}{F_{G,\max} - F_{G,\max}^{\text{nomal}}} \times F_G + \frac{\bar{F}_g \times F_{G,\max} - F_{G,\max}^{\text{nomal}}}{F_{G,\max} - F_{G,\max}^{\text{nomal}}} \quad (11)$$

$$\Delta F(F_G) = \frac{\bar{F}_g - 1}{F_{G,\min}^{\text{nomal}} - F_{G,\min}} + \frac{F_{G,\min}^{\text{nomal}} - \bar{F}_g \times F_{G,\min}}{F_{G,\min}^{\text{nomal}} - F_{G,\min}} \quad (12)$$

Considering the influence of real-time operation conditions on reliability model, it is not difficult to understand the frequent occurrence of chain failures in real-time systems. According to the definition of time-varying reliability model of power system, after the initial failure, the operating conditions of the system change. Some elements are running near the limit value. At this time, the outage probability of these components is not statistical average but close to 1, and new faults are likely to occur. So the probability of chain failure is the product of single outage probability based on real-time operation conditions, which is much greater than the product of single fault probability based on statistical average defined by traditional reliability evaluation method.

## 2.2 Power System Operation State Evaluation Algorithm

The key of power grid operation reliability evaluation is to solve the problem of reliability model of components under operation conditions, especially in fault state and special operation mode. As an important component of power grid, the reliability level of generator set is closely related to the system reliability [9]. In the actual operation of power grid, the main factors that affect the operation of generating units include start-up time, start-up failure probability, load increase rate and outage delay.

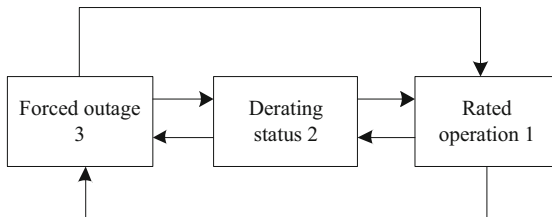
At the same time, the operation modes of units are different when they are put into operation at different locations, such as base load or peak shaving operation, and under

different standby modes, such as cold standby, hot standby and rotary standby. Therefore, it is necessary to further study the generator unit model suitable for operation planning reliability evaluation [10]. It takes a certain time for each unit from receiving the order to put into operation to full output state, which is called start-up time, acceleration time or start-up lead time. According to the type of unit and the state, the time can be divided into several minutes to more than ten hours. The start-up delay characteristic will delay the rapid recovery of the accident, prolong the fault duration and increase the system risk. Based on this, the startup delay of typical units in power system is investigated and recorded, as shown in the table below (Table 1).

**Table 1.** Startup delay of typical units in power system

Unit type	Time delay $T_S$ (min)	
	Hot state	Cold state
Gas turbine	3 ~ 5	10 ~ 15
Condensing thermal power unit	30 ~ 60	120 ~ 180
Hydropower and pumped storage power stations	0.5 ~ 2	1 ~ 3
Variable pressure operation thermal power unit	60 ~ 80	360
Extraction type thermal power unit	30 ~ 60	90 ~ 180

Based on the information in the above table, two state or three state models are further used to describe the reliability of the above units. The specific model is shown in the figure below (Fig. 4):



**Fig. 4.** Simulation diagram of three state unit in power system

For intermittent operating units, all states of the generator can be divided into: the set of forced shutdown states when the system is needed, the set of forced shutdown states when the system is not needed, and the set of running states. According to the definition of CFOR, the general calculation can be obtained formula:

$$CFOR = \frac{\sum_{i \in I} p_i}{\sum_{i \in I} p_i + \sum_{i \in M} p_i} \tag{13}$$

The probabilistic reasoning of system reliability is based on the k-nearest neighbor technology, which is used to calculate the probability interval of each state of the node random variable under the condition of known correlation and evidence. The reasoning method includes precise reasoning algorithm and approximate reasoning algorithm. The precise reasoning algorithm of reliability. The joint probability mass function algorithm formed by traversing the vertices of the reliability set is:

$$\bar{P}(x_q | \mathbf{x}_e) = \Delta F(F_G) \max_{j=1, \dots, N} \frac{\sum_{X_{N1}} \prod_{i=1}^N P_j(x_i | \pi_i)}{\sum_{X_{M2}} \prod_{i=1}^N P_j(x_i | \pi_i)} \tag{14}$$

Based on this further calculation, before the imprecise condition estimation of the power system outage probability, the mathematical expression of the equipment outage probability is clarified, and the relationship between the outage rate and the outage probability is clarified. Reliability refers to the probability that the equipment will operate normally within time t, that is, the probability that the equipment will not be out of service within time t. Suppose the continuous random variable T (T > 0) represents the time from continuous normal operation of the equipment to before shutdown, then the reliability function can be expressed as:

$$R(t) = P - \bar{P}(x_q | \mathbf{x}_e)(T \geq t) \tag{15}$$

The fault distribution function F(t) describes the probability of a power equipment outage event before the arrival of the equipment at time t. It can be seen that its relationship with the reliability function is:

$$F(t) = 1 - R(t) \tag{16}$$

Outage rate is a commonly used reliability index. Its statistical description is the ratio of equipment outages per unit time. It is defined by the following formula, and its probability is described as:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{R(t) - R(t + \Delta t)}{F(t)\Delta t \cdot R(t)} = \frac{dF(t)}{F(t)dt \cdot (1 - F(t))} \tag{17}$$

When the transmission line is under heavy load operation, as its carrying capacity increases and reaches the set value, it will trigger the relay protection action and cause it to stop. Therefore, the probability of line outage under this condition will also increase. Therefore, the operating conditions that significantly affect the reliability of the transmission line and the operating status of the transmission line are used as the node evidence variables of reliability, and the reliability estimation model of the power system operating state reliability estimation is constructed, as shown in the Fig. 5.

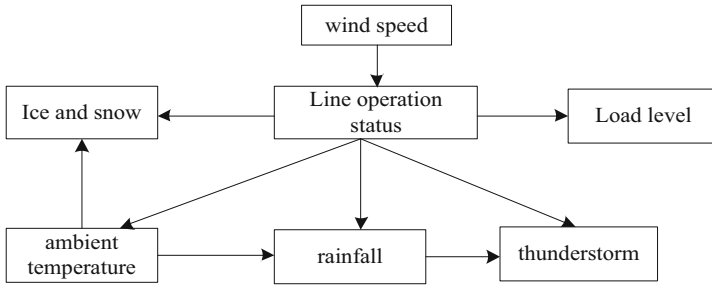


Fig. 5. Reliability estimation reliability of power system operating state

In the figure, in order to realize the inaccurate condition estimation of the outage probability of a transmission line under a given operating condition, it is necessary to know the prior probability of the equipment operating state and the probability correlation between the line operating state and the operating condition. The prior probability of line outage can be replaced by the long-term statistical outage probability of the line. For the probabilistic correlation between the line operating state and its operating conditions, the condition  $K(E_a|H_z), a \in \{1, 2, 5\}, z \in \{1, 2\}, K(E_3|H_z, E_{1,b1})$  corresponding to each evidence node is used to analyze the reliability of the power system based on a deterministic method, that is, for certain expected system operation modes or events, through power flow calculation or stability Performance analysis, to clearly obtain these system operation modes or the consequences of the system after the occurrence of the event, in order to determine the reliability of the power system. The decision-making goal based on the deterministic criterion is generally that the power system will not fail under these expected methods. By introducing the K nearest neighbor theory, we not only analyze the results of each scene, but also consider the possibility of each scene, and take the characteristics of the K nearest neighbor theory to characterize the reliability of the system. When the number of scenes is large enough, You can get the probability method of power system reliability analysis and risk assessment as shown in the formula. The simulation method to calculate the reliability of the power system can be uniformly expressed in the form shown in the formula.

$$E[f] = \sum_{x \in X} f(x) \cdot \frac{n_x}{\lambda(t)n} = \frac{1}{\lambda(t)n} \sum_{x=1}^n f(x) \tag{18}$$

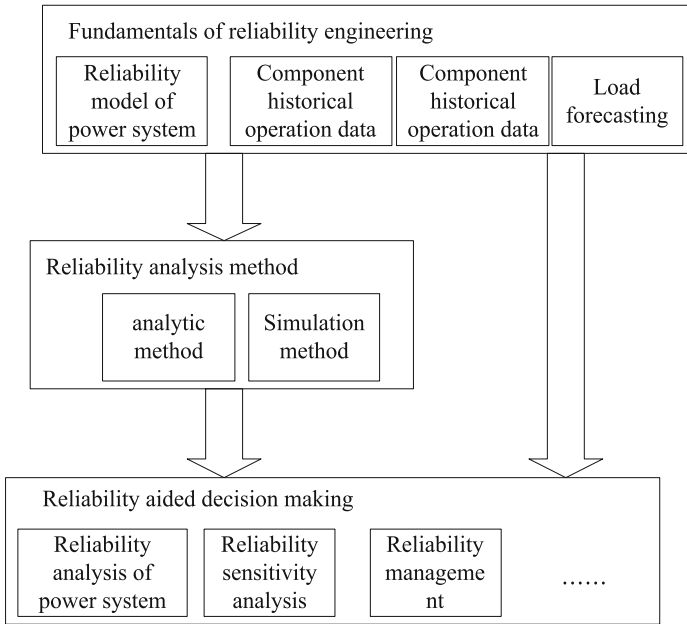
Under normal circumstances, “year” is the main unit of reliability analysis, which conducts a comprehensive analysis of the average reliability of various operating states of the system. The main indicators include two indicators: system and load. The system indicators mainly include indicators such as the time and frequency of the system’s continuous power outages, lines, etc., while the load indicators include the duration of power outages and the average failure probability of the load point. Evaluating the



reliability can effectively improve the planning and design system. Therefore, the operation reliability of the distribution network must show the load loss and load margin, and use this as the basic condition to show the reliability of the entire system. The operational reliability index mechanism accommodates many index variables. At the same time, it is difficult to model or calculate. In addition, there is a lot of complex information in each index, and analysis one by one will shoulder a huge workload.. Among them, the principal component analysis should adopt the multivariate statistical analysis method, which can make the modeling easier, and can also extract valuable information from it, and reduce the variable dimension as much as possible. Principal component analysis generally uses variance, sum of squares of deviations, etc. to perform calculations on various indicators, clean up repeated indicators, and extract indicators that have little relevance and include a lot of original information. Afterwards, the components and ranges that will interfere with the system's operational reliability are selected in a targeted manner, so as to gradually reduce the evaluation range. Standardized evaluation indicators. Under normal circumstances, various operational feasibility indicators have obvious differences, and they are distributed in different places, and they must be processed according to the indicator parameters. Therefore, it is necessary to use distribution big data to obtain various index distribution functions, carry out normal distribution according to various index variables, and then use this as a standard to transform into normal distribution variables. Regarding matrix creation and eigenvalue calculations, the values of other variables are established.

### **2.3 Realization of Reliability Analysis of Power System Operation State**

Security can also be called dynamic reliability, which refers to the ability of the power system to withstand the disturbance after scene switching, such as whether the power system can return to its original operating state or transition to a state after a sudden short circuit or loss of system components. The new stable operating state, the ability to provide users with electrical energy uninterrupted. The basic evaluation index for the reliability of the power system's operating state is: uninterrupted supply of qualified power to all users. The essence of power system reliability and risk research is to pre-consider the probabilities and consequences of various operating modes, and make comprehensive decisions to give full play to the potential of each device in the system, so as to meet the load requirements of all users with quality and quantity. Based on this, the main frame structure of power system engineering reliability is shown in the following Fig. 6:



**Fig. 6.** Power system engineering reliability model

The main task of the reliability model of power system engineering is to accumulate historical operating data of components and component reliability test data, and analyze component reliability models and parameters. Based on the power system reliability model and expected load changes, analyze or simulate the phenomenon that the power system cannot complete the specified function within the specified time, and calculate its occurrence probability and consequences to obtain quantitative evaluation indicators and standards. On the basis of coordinating reliability and system input, comprehensive evaluation and auxiliary decision-making of power system operation and control are carried out. Find out the key links that restrict the reliability of the power system, and propose specific measures to improve and increase the reliability level, and organize or assist relevant departments to implement them. Power system reliability analysis refers to the probability analysis and evaluation of the power system's ability to continuously and uninterruptedly supply power that meets the voltage quality requirements of the regulations to each load point within a specified time and under specified conditions by using reliability indicators, and find out the impact the key link of system reliability. Reliability analysis and risk assessment of the power system in the operating environment are based on the effective information in the operating environment, including deterministic and predictive information, to establish component reliability models, and to consider the recent possible operating status and random failures of the power system, analyze and predict The adequacy and safety of short-term operation of the power system. The focus of power system reliability analysis and evaluation is different, and the selected reliability indicators are not the same, but they can all be

expressed as the expected value of a certain measurement function, as shown in the formula.

$$\Delta E[f] = E[f] / \sum_{x \in X} f(x)p(x) \tag{19}$$

The sufficiency evaluation of wiring reliability is mainly based on the ability of wiring to distribute load demand. Traditional adequacy indicators include the following four categories: expected value, probability, frequency and duration. Based on the non-sequential simulation method, the following indicators are selected to measure the reliability level of plant and station wiring:

$$EENS = \frac{\sum_{k=1}^n (P_{bk} \cdot t_k)}{\Delta E[f] - n} \tag{20}$$

Since the sufficiency analysis only involves the steady-state calculation of the power system, the analysis models and calculation methods are relatively mature, and the requirements for calculation time are not high, so the existing research and

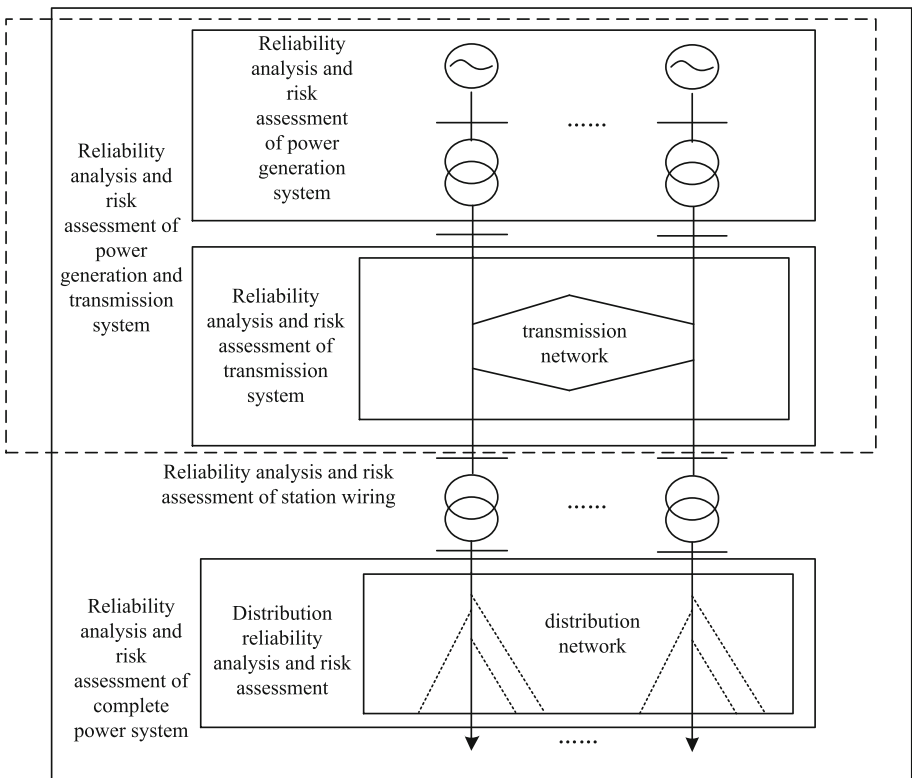


Fig. 7. Hierarchical levels of power system reliability analysis

application of power system reliability analysis and risk assessment are mainly concentrated on This aspect. With the increase in the number of components in the power system to be analyzed, the number of system scenarios  $X$  has also expanded dramatically. Due to the limitations of calculation tools and calculation methods, it is still difficult to calculate and analyze the reliability and risk of the entire power system in a unified manner.. According to different functions, structures, voltage levels, etc., the power system can be divided into different subsystems such as power generation, transmission, and distribution in the project. The wiring form, operation mode, equipment redundancy, failure loss, etc. of each subsystem are different, and the strategies used in the calculation of their sufficiency are also different. Therefore, the reliability of these subsystems is usually analyzed separately and based on this combination produces different levels of power system reliability analysis, as shown in the Fig. 7.

As a key hub in the power grid, the reliability of the substation plays a vital role in the reliability of the entire power grid. Except for transformers, the electrical distances of other components in the plant and station wiring are relatively short and the cross-section selection is relatively large. They are all regarded as non-impedance components in the power system analysis process, and only need to be connected to judge to generate equivalent nodes. No need to directly participate in the power flow calculation. In addition to the limited number of internal components in the plant and the large wiring redundancy, the reliability of the plant's wiring is generally independently studied to obtain an equivalent bus model with corresponding probability characteristics. When performing system reliability analysis, this model is generally used directly, instead of unfolding the plant and station wiring. Rely on the parallel relationship to mine the main influencing factors. At present, there are many factors that affect the reliability of the distribution network. Therefore, it is difficult to establish an accurate prediction model, and the speed accuracy will also be affected. If you want to improve this situation, you can use the association rule mining method to extract the influencing factors from the heterogeneous multi-source data of the distribution network, and bring a certain reference basis for the establishment of the prediction model, reduce the input dimension as much as possible, and further Improve the prediction speed, thereby improving the reliability of the power system.

### 3 Analysis of Results

In order to verify the necessity and effectiveness of the power system real-time operating state reliability method based on  $K$  nearest neighbors, a comparison experiment was carried out with double busbars with bypass and 3/2 wiring as examples. For the sake of comparison, there are two incoming lines and four outgoing lines for the two types of wiring. The power upper limit of the incoming line is 250 kW, and the power demand of the outgoing line is 100 kW. To ensure the true and effective experimental results, the specific experimental equipment and parameters are unified Settings, as shown in the following table (Table 2):

**Table 2.** Experimental equipment and parameters

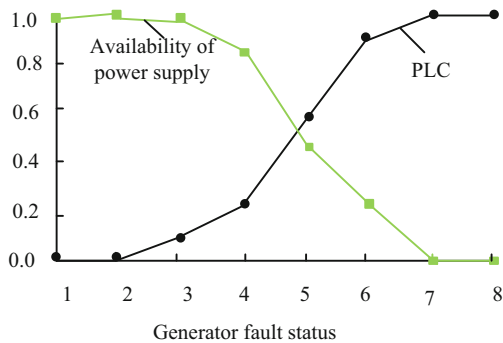
Component type	Average repair time $Y_i$ /(H/time)	Unreliability $q_i$ /(times/(set))
High voltage circuit breaker	160	0.06
High voltage disconnecter	56	0.00149
High voltage bus	20	0.015

Further regulate the operating conditions of the power system, as shown in the table (Table 3).

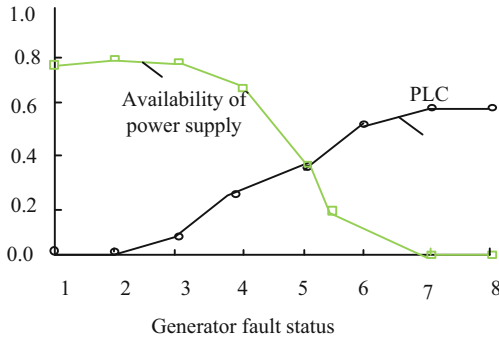
**Table 3.** Power system operating conditions

state	Generator fault description	Generator voltage/pu	Load voltage/pu	System frequency/Hz	Line transmission capacity/MW	Reserve capacity MW
1	0 fault	1.010	1.019	50.00	168.5	415.7
2	1 fault	1.006	1.007	49.73	168.4	125.7
3	2 fault	1.001	0.9879	49.15	168.4	15.7
4	2 faults + 1 unit output reduction 10 MW	1.000	0.9873	48.64	168.4	5.7
5	2 faults + 1 unit reduced output 30 MW	0.9997	0.9857	47.57	168.4	0.0
6	2-fold fault + 1 set of reduction output 40 MW	0.9994	0.9848	47.02	168.4	0.0
7	2-fold fault + 1 set of reduction output 50 MW	0.9989	0.9839	46.44	168.4	0.0

Based on the above environment, the reliability index of the power system in the fault state is further detected, as follows (Figs. 8 and 9):



**Fig. 8.** The reliability index of the method in this paper under the failure state



**Fig. 9.** Reliability index under the failure state of the traditional method

According to the evaluation results, the following conclusions can be drawn: the reliability level of the traditional system under the operating state is much lower than the reliability level under the operating state of this article, and the traditional reliability evaluation cannot express the reliability of the system under the generator failure state. Level. Based on this, further statistics are made on the outage probability during the operation of the power system and the reliability index under the power system line fault state, as shown in the table (Tables 4 and 5).

**Table 4.** Probability of component outage in power system fault state

State	Generator outage probability	Line outage probability	Load outage probability
1	$1.668 \times 10^{-2}$	1.000	0
2	$1.668 \times 10^{-2}$	$8.092 \times 10^{-1}$	0
3	$1.668 \times 10^{-2}$	$1.811 \times 10^{-4}$	0
4	$1.668 \times 10^{-2}$	$1.811 \times 10^{-4}$	0
5	–	–	–

**Table 5.** Reliability index under power system line fault state

State	PLC	SI/(system minute/year)	EEMS/(MWh/a)	Availability of power supply
1	1.0000	$5.256 \times 10^5$	$8.410 \times 10^6$	0.0000
2	1.0000	$5.256 \times 10^5$	$8.410 \times 10^6$	0.0000
3	0.8819	$4.075 \times 10^5$	$6.517 \times 10^6$	0.2247
4	0.0011	$7.714 \times 10$	$1.234 \times 10^3$	0.9998
5	0.0002	$1.596 \times 10$	$2.554 \times 10^2$	0.9999

According to the evaluation results, the following conclusions can be drawn: the reliability level of the system in the fault state is much lower than the reliability level in the normal operation state, and the traditional reliability evaluation cannot express the reliability level of the system in the line fault state. In order to further test the effectiveness of the method proposed in this article, the simulation system generates a total

of 50,000 h of samples, and each sample contains the operating condition information and equipment operating status information during the period. Compare the standard reliability values during the operation of the power system for comparative testing to determine the authenticity of the state reliability test results. The specific test results are shown in the following Fig. 10:

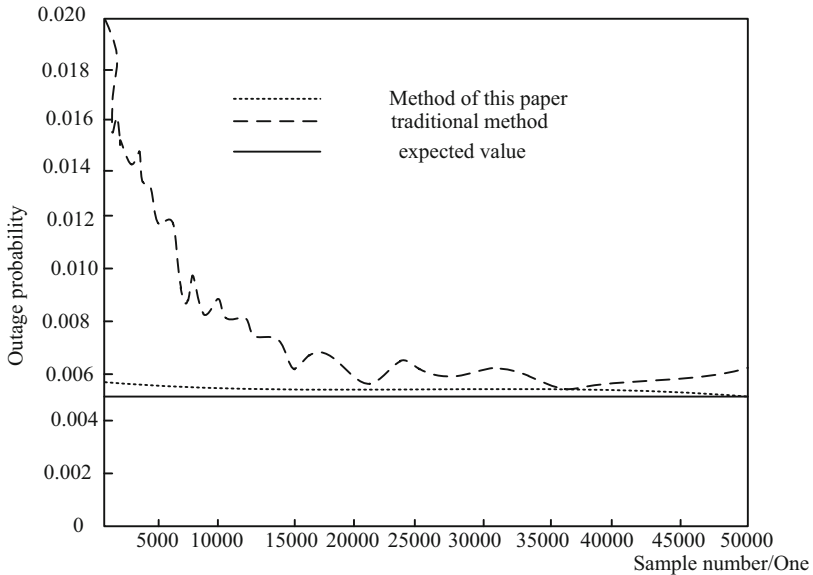


Fig. 10. Comparison test results

It can be seen from the figure that the method proposed in this paper has advantages over traditional methods, and is basically consistent with the expected operating state reliability curve. It can be seen from the figure that the estimation result obtained by the method in this paper is always significantly close to the true value of the line outage probability, and as the number of samples increases, the outage probability interval obtained by the method in this paper gradually shrinks and converges to the true line outage probability. Probability reflects the reasonableness of the estimation results of this paper.

#### 4 Concluding Remarks

The development of science and technology makes the power system more and more perfect. In the context of big data, big data technology has been widely used, which has allowed the power industry to develop further. At present, because many factors will affect the reliability of the operation of the distribution network, relevant workers must analyze the relevant influencing factors in an all-round way when evaluating the reliability of the operation, and formulate a complete control plan to ensure the

distribution the net can get sustainable development. Based on this, an optimization study on the reliability analysis method of power system real-time operation state based on K nearest neighbors is carried out, so as to better guarantee the power system operation quality. In addition, it is necessary to use the big data of the distribution network correctly to extract valuable data from a large amount of information, so as to improve the reliability of the operation of the distribution network.

**Acknowledgement.** This paper is supported by the “Research and demonstration application of key technologies of intelligent control robot assistant (SGTYHT/19-JS-215)”, the project of headquarters management science and technology of State Grid Corporation of China.

## References

1. Liao, M., Chakraborty, A.: Optimization algorithms for catching data manipulators in power system estimation loops. *IEEE Trans. Control Syst. Technol.* **27**(3), 1203–1218 (2019)
2. Kumar, R., Sharma, V.K.: Whale optimization controller for load frequency control of a two-area multi-source deregulated power system. *Int. J. Fuzzy Syst.* **22**(1), 122–137 (2020)
3. Chaubey, P., Lather, J.S., Yeliseti, S., et al.: Robust power system stabilizer based on static output feedback approach to enhance power system stability. *Energy Procedia* **158**(2), 2960–2965 (2019)
4. Seck, G.S., Krakowski, V., Assoumou, E., et al.: Embedding power system’s reliability within a long-term energy system optimization model: linking high renewable energy integration and future grid stability for France by 2050. *Appl. Energy* **257**(11), 114037–114039 (2020)
5. Jimada-Ojuolape, B., Teh, J.: Impact of the integration of information and communication technology on power system reliability: a review. *IEEE Access* **8**(10), 24600–24615 (2020)
6. Peyghami, S., Davari, P., Blaabjerg, F.: System-level reliability-oriented power sharing strategy for DC power systems. *IEEE Trans. Ind. Appl.* **55**, 4865–4875 (2019)
7. da Costa, L., Thomé, F.S., Garcia, J.D., et al.: Reliability-constrained power system expansion planning: a stochastic risk-averse optimization approach. *IEEE Trans. Power Syst.* **39**, 97–106 (2020)
8. Zhao, Y., Xie, W., Jiang, J., et al.: Replacement of Marx generator by tesla transformer for pulsed power system reliability improvement. *IEEE Trans. Plasma Sci.* **47**(1), 574–580 (2019)
9. Abunima, H., Teh, J., Jabir, H.J.: A new solar radiation model for a power system reliability study. *IEEE Access* **7**, 64758–64766 (2019)
10. Altamimi, A., Jayaweera, D.: Reliability of power systems with climate change impacts on hierarchical levels of PV systems. *Electr. Power Syst. Res.* **190**, 116–123 (2021)





# Low Resolution License Plate Recognition Based on Intelligent Data Processing and Prediction Algorithm

Mi Meng<sup>1</sup>(✉) and Chun-hu He<sup>2</sup>

<sup>1</sup> Chongqing Telecommunication Vocational College,  
Chongqing 402274, China  
mengmi4545@163.com

<sup>2</sup> Chongqing Toramon Technology Co., Ltd., Chongqing 401121, China

**Abstract.** Aiming at the problems of low accuracy of low resolution license plate recognition and long time consuming of recognition and registration in traditional methods, this paper proposes a low resolution license plate recognition method based on intelligent data processing and prediction algorithm. Firstly, the low resolution license plate is located, and the low resolution digital image of license plate is defined by the principle of image registration; Secondly, the doc scale space is constructed to determine the Gaussian pyramid and Gaussian difference pyramid model of low resolution license plate, and the RANSAC prediction algorithm is used to eliminate the mismatching of low resolution license plate and realize low resolution license plate recognition. The experimental results show that the proposed method can achieve fast recognition accuracy and recall rate of low resolution license plate, and the recognition time is low.

**Keywords:** RANSAC prediction algorithm · Low resolution license plate recognition · Image registration · Gauss pyramid · Gauss difference pyramid

## 1 Introduction

The research on license plate recognition technology abroad began in the 1980s. At that time, it was usually only used some of the simplest image processing technologies to process a module of license plate recognition process, without a complete process system, and no real-time automation. In 1990, lotufo and others developed a license plate recognition system, which consists of three parts: extracting character features, building character template library, character segmentation and character template matching, which was of epoch-making significance at that time. But after the 1990s, the complete license plate recognition system was formed slowly, all of which was due to the rapid development of computer vision and image processing technology [1–3]. China's license plate recognition technology is from the 1990s, due to the particularity of Chinese characters in China's license plate, can not fully learn from foreign technology, so the gradual rise of the domestic license plate recognition system. At present,

the better license plate recognition systems include Hanwang eye developed by Hanwang company, license plate communication developed by Automation Department of Zhejiang University, Institute of automation of Chinese Academy of Sciences, and national key laboratories in domestic universities. They are all studying license plate recognition related algorithms and have achieved good results [4, 5].

Generally, when the car body color is different from the license plate color, the license plate detection method based on color is better. Relevant scholars use HSI color model to detect the candidate area of license plate, and then verify it by position histogram. The image is divided into different regions according to different colors by means of mean shift algorithm, and then the license plate is distinguished according to the characteristics of rectangle, aspect ratio and edge density. Domestic scholars propose to first generate the gray image of the license plate, remove the non license plate area according to the color accompanying attributes of the character area and the background area, retain the real license plate, and then use the color information to get more accurate positioning results [6]. Color based method can detect slanting or deformed license plate, but it can't distinguish other objects with similar color and size, and is very sensitive to various lighting changes.

In order to solve the above problems, this paper proposes a low resolution license plate recognition method based on intelligent data processing and prediction algorithm.

## **2 Low Resolution License Plate Location**

### **2.1 Principle of Image Registration**

License plate location can also be called license plate detection. The existing methods can be divided into two categories: traditional license plate location method and license plate location method based on deep learning. Traditional license plate location methods can be divided into texture features, edge features, character region features and color features; The deep learning method is mainly based on the deep learning detection network. Because the license plate is a rectangular region with a specific aspect ratio, and its edge density is richer than other positions in the image, the edge information is widely used for license plate detection. Generally, the edge detector and some morphological operations are combined to generate the candidate license plate rectangular region. Then Hough transform is used to find the vertical and horizontal lines as the license plate boundary. Although edge based methods are fast in detection speed, they are not suitable for complex or blurred license plate images. They are too sensitive to useless edges in images [7–9]. Image registration is the best matching process of two or more images obtained by different imaging equipment and under different conditions, such as climate, camera angle and position. That is to find out the relative position parameters such as scaling factor, relative rotation angle and relative displacement between these images, and then use the method of geometric transformation to classify these images into a unified coordinate system.

## 2.2 Definition of License Plate Low Resolution Digital Image

Sample a simulated license plate image, set up a data, arrange these data into a data matrix according to the relative position of the sampling points, and then quantify the amplitude of each element, so as to obtain the digital matrix of the license plate image. The elements of the matrix are called pixels or pixels [10, 11].

Let the coordinate value of the origin be  $(1, 1)$ , then an  $M \times N$  size digital image,  $f(x, y)$  can be represented by a matrix:

$$f(x, y) = \begin{bmatrix} f(1, 1) & f(1, 2) & \cdots & f(1, N) \\ f(2, 1) & f(2, 2) & \cdots & f(2, N) \\ \cdots & \cdots & \cdots & \cdots \\ f(M, 1) & f(M, 2) & \cdots & f(M, N) \end{bmatrix} \quad (1)$$

It can be seen from the above that digital image  $I_1(x, y)$  can be represented by a two-dimensional matrix, let  $I_1(x, y)$  represent the gray value of the image to be registered at point  $(x, y)$ , and let  $I_2(x, y)$  represent the gray value of the reference image at point  $(x, y)$ , then the registration relationship between the reference image and the image to be registered is expressed as follows:

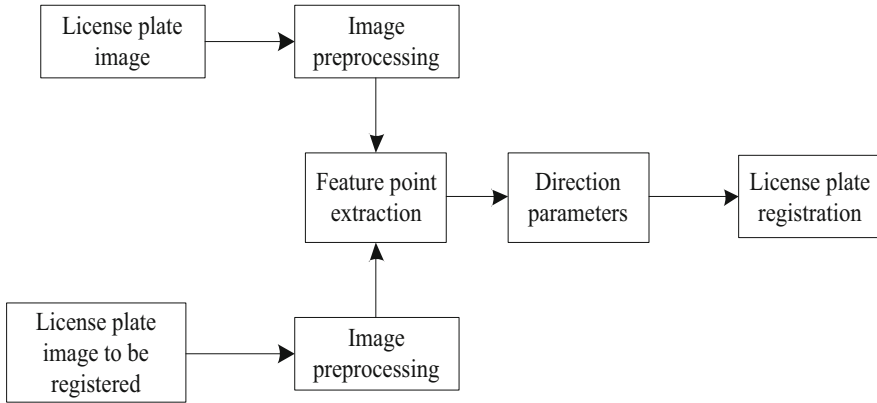
$$I_2(x, y) = g(I_1(f(x, y))) \quad (2)$$

In formula (2),  $f$  is a two-dimensional geometric transformation function and  $g$  is a one-dimensional gray scale transformation function. The coordinate mapping relationship between reference image and image to be registered is established. Image registration can be considered as the matching of two images in gray and spatial geometry. So the main purpose of image registration is to find the best gray transformation relationship and spatial transformation relationship.

## 3 Low Resolution License Plate Recognition Based on Intelligent Data Processing and Prediction Algorithm

### 3.1 License Plate Image Registration Algorithm Based on Intelligent Data Processing

The full name of sift is scale invariant feature transformation one, which is a feature matching operator based on scale space and keeping the scale and rotation of license plate image unchanged. The local features of the image are extracted by the registration algorithm, which has good anti noise performance and fast processing speed. Therefore, in the field of registration, the registration algorithm is a common registration algorithm [12, 13]. Registration algorithm includes three aspects: content feature detection, feature description and feature matching (Fig. 1):



**Fig. 1.** Steps of license plate image registration algorithm

In this paper, the feature registration algorithm is used. Before two images are registered, the feature points of the two images must be found first, and these feature points are the extreme points in the scale space of the image, so the scale space of the image must be constructed first [14–16].

The principle of scale space. The concept of scale space was first proposed in the field of computer vision, mainly to describe the scale characteristics of license plate image. By proving that the transform kernel of scaling transform is Gaussian convolution kernel, and it is unique. Later, Gaussian convolution kernel is proved to be linear kernel, and it is unique. The scale space of a two-dimensional image is obtained by convolution of the image and Gaussian function. Suppose that the two-dimensional image of license plate is represented by  $I(x, y)$ , and  $G(x, y; \sigma)$  is the Gaussian function, then the scale space of the two-dimensional image is represented by:

$$L(x, y; \sigma) = I(x, y) * G(x, y; \sigma) \tag{3}$$

Gauss function  $G(x, y; \sigma)$  is expressed as:

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{4}$$

In formula (3),  $L(x, y; \sigma)$  is the scale space obtained by Gaussian convolution,  $\sigma$  is the scale factor, that is, the variance of Gaussian function distribution, and the value of  $\sigma$  directly affects whether an appropriate image scale space can be established. It can be seen from Eq. (3) that when  $\sigma = 0$  is set:

$$L(x, y; 0) = I(x, y) * G(x, y; 0) = I(x, y) \tag{5}$$

At this time,  $G(x, y; 0)$  is a two-dimensional pulse function. Therefore, the value of  $\sigma$  can not be too large. The image data described in this way is too rough to reflect the

details of the image. The value of  $\sigma$  can not be too small. The image data described in this way is too fine to reflect the image information completely.

### 3.2 Establishment of DOC Scale Space

In order to extract more stable feature points of license plate image, the Gaussian difference scale space model doc is constructed based on the Gaussian scale space model. Take different values for  $\sigma$ , and then according to Eq. (3),  $I(x, y)$  and  $\sigma$  are different. By convoluting the Gaussian function of  $G(x, y; \sigma)$ , the Gaussian pyramid model of the image can be obtained, and more feature points of the license plate image can be extracted from different scale spaces. In general, the Gaussian pyramid model has order, and each order takes the layer image. In the Gaussian pyramid model, the scale space of DOC can be obtained by subtracting the functions of two adjacent scale spaces.

$$D(x, y; \sigma) = I(x, y) * (G(x, y; k\sigma) - G(x, y; \sigma)) = L(x, y; k\sigma) - L(x, y; \sigma) \quad (6)$$

In Eq. (6),  $D(x, y; \sigma)$  is the Gauss difference function. In the Gaussian pyramid model, there are layers of images in each order, and the scale difference between each layer is a constant  $k$  times. By subtracting two adjacent layers of images,  $D(x, y; \sigma)$  can be obtained. Thus, the model is formed, as shown in Fig. 2.

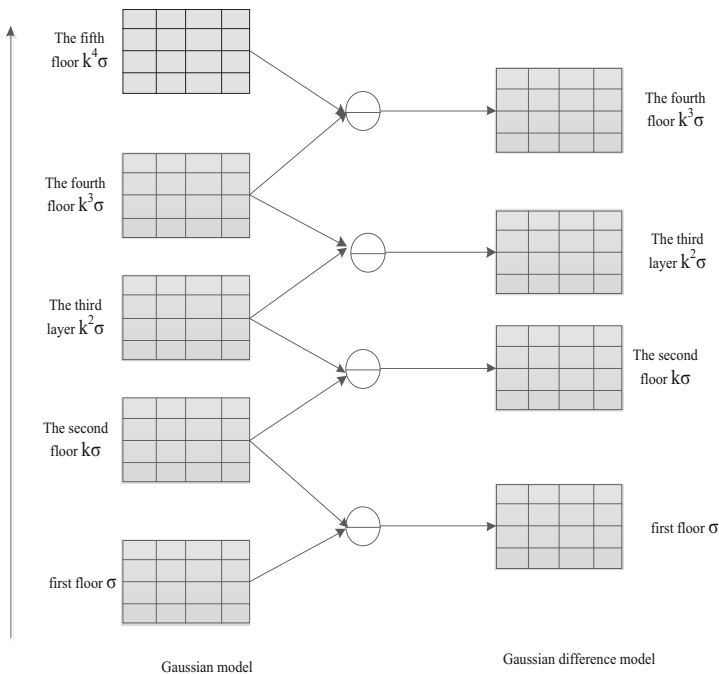


Fig. 2. Gaussian pyramid and Gaussian difference pyramid

### 3.3 Eliminating Mismatching Based on RANSAC Prediction Algorithm

The parameters of the mathematical model of RANSAC algorithm are estimated by iterative method. When the error rate of data exceeds, the algorithm can also deal with it. Moreover, the model is relatively simple and has strong adaptability. It is one of the most robust algorithms. Because of these advantages, it is often used in the field of registration. In this paper, the algorithm is added to the original method to better eliminate the mismatched point pairs and improve the accuracy of license plate image registration.

The idea of RANSAC algorithm is simple and ingenious. Its input value is the collection of measurement data, which is called data point. In the data point set, some data points can satisfy the pattern of some unknown parameters. These data points are called interior points, and all data points except interior points are exterior points. The set of matching points is sampled for many times, and a set of allowable error range is set. All the points are in the range and match with the model, while most of the outer points fall outside the allowable error range, so the model determined by the inner point is close to the actual model, so the parameters of the model can be determined. RANSAC algorithm is to carry out many random sampling experiments, eliminate all the feature points that are not in the error range, and optimize the parameters of the model.

In this paper, based on the Euclidean distance method, the algorithm is added to better eliminate the mismatching point pairs and improve the accuracy of license plate image registration:

- (1) The threshold value of Euclidean distance method is given a  $\delta$  value, the range of which is selected from one to another, and the  $\delta$  value is adjusted according to the needs of the experiment. If the value of NN/SCN is smaller than the set closed value, the point is retained, otherwise the point is deleted.
- (2) In the last step, among the remaining pairs of matching points, randomly sampled pairs of matching points are obtained as samples of data points.
- (3) The transformation matrix  $H$  is calculated by using the  $m$  matching points;
- (4) According to the transformation matrix, the error values of the remaining pairs of matching points are calculated. Let  $(x, x')$  be the feature points in two images corresponding to the same position, that is, a pair of matching points. If  $\delta$  is a threshold, then there is a threshold:

$$\|Hx - x'\|^2 \leq \delta \quad (7)$$

- (5) Judge whether (7) is true. If it is true, then the corresponding feature points are interior points, and the number of interior points is  $n$ . If it is not true, then the corresponding feature points are exterior points and are eliminated.
- (6) The new transformation matrix  $H$  is solved by  $n$  interior points in, and the new number of interior points  $n'$  is obtained by performing (4) and (5).
- (7) Judge the size of  $n$  and  $n'$ . If  $n' > n$ , then repeat (2)–(7). If two values are equal, then the number of interior points tends to be stable, and exit ends.

After repeated execution for many times, more than 90% of the outliers can be removed, which improves the accuracy of RANSAC algorithm to solve the geometric transformation model of license plate image.

## 4 Experiment

In order to verify the practical application effect of the proposed low resolution license plate recognition method based on intelligent data processing and prediction algorithm, the simulation experiment is carried out. The image data used in this experiment comes from MySQL database, and 500 images are selected from the database for the experiment. Due to the space limitation of the experiment, only one image is displayed (Fig. 3).



Fig. 3. Experimental image

### 4.1 Experimental Design

At present, there is no unified index to evaluate the performance of license plate location in various natural scenes. In this paper, we use the accuracy (P) and recall (R) to quantify the results of license plate location. The accuracy is defined as the ratio of the number of license plates correctly detected to the total number of candidate regions generated. The more candidate regions generated by the location network, the lower the accuracy of the model.

$$P = \frac{T_p}{T_p + F_p} = \frac{T_p}{N} \quad (8)$$

where  $T_p$  is defined as the total number of candidate regions for the correct detection of license plate,  $F_p$  is the total number of candidate regions for the wrong detection of license plate, and  $N$  is all candidate regions generated by the network.

Recall rate is defined as the ratio of the number of license plates correctly detected to the total number of real frames marked by the license plate. The specific formula is as follows

$$R = \frac{T_p}{T_p + F_n} \quad (9)$$

where  $F_n$  is the total number of undetected license plate label real boxes. The more the number of license plates detected correctly, the higher the recall rate.

## 4.2 Experimental Analysis

### Low Resolution License Plate Recognition Accuracy

In order to detect the accuracy of low resolution license plate recognition, the paper adopts the methods of literature [7], literature [8], literature [9] and the method in this paper to test the accuracy of low resolution license plate recognition. The results are shown in Table 1.

**Table 1.** Accuracy of low resolution license plate recognition under different methods

Iterations/time	Low resolution license plate recognition accuracy/%			
	Methods in literature [7]	Methods in literature [8]	Methods in literature [9]	Method of this paper
5	76	66	72	99
10	75	69	54	96
15	54	73	59	89
20	63	75	64	94
25	73	79	67	96
30	70	72	63	93
Mean value	68.5	72.3	63.2	94.5

Analysis of Table 1 shows that there are differences in the accuracy of low resolution license plate recognition under different methods. When the number of iterations is 5, the low resolution license plate recognition accuracy of literature [7] is 76%, the low resolution license plate recognition accuracy of literature [8] is 66%, the low resolution license plate recognition accuracy of literature [9] is 72%, and the low resolution license plate recognition accuracy of this method is 99%. When the number of iterations is 20, the low resolution license plate recognition accuracy of the method in literature [7] is 63%, the low resolution license plate recognition accuracy of the method in literature [8] is 75%, the low resolution license plate recognition accuracy of the method in literature [9] is 64%, and the low resolution license plate recognition accuracy of the method in this paper is 94%. The low resolution license plate recognition accuracy of this method is much higher than other methods.



### Time for Low Resolution License Plate Recognition

In order to verify the low resolution license plate recognition efficiency of this method, the time-consuming detection experiments of low resolution license plate recognition are carried out by using the methods of literature [7], literature [8], literature [9] and this method, and the results are shown in Table 2.

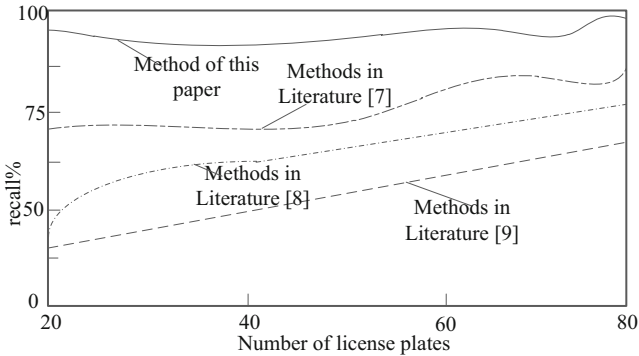
**Table 2.** Time consumption of low resolution license plate recognition under different methods/s

Iterations/time	Time for low resolution license plate recognition/s			
	Methods in literature [7]	Methods in literature [8]	Methods in literature [9]	Method of this paper
5	35	62	76	4
10	42	67	74	6
15	52	60	72	5
20	46	57	73	7
25	55	52	75	5
30	58	49	70	7
Mean value	48	57.8	73.3	5.7

According to the analysis of Table 2, the time of low resolution license plate recognition is different under different methods. When the number of iterations is 10, the low resolution license plate recognition time of reference [7] is 42 s, the low resolution license plate recognition time of reference [8] is 67 s, the low resolution license plate recognition time of reference [9] is 74 s, and the low resolution license plate recognition time of this method is 6 s. When the number of iterations is 30, the low resolution license plate recognition time of reference [7] method is 58 s, the low resolution license plate recognition time of reference [8] method is 49 s, the low resolution license plate recognition time of reference [9] method is 70 s, and the low resolution license plate recognition time of this method is 7 s. The low resolution license plate recognition time of this method is far less than the other two methods.

### 4.3 Recall Rate of License Plate in Different Ways

In order to verify the efficiency of low resolution license plate recognition in this paper, the paper adopts the methods of literature [7], literature [8], literature [9] and this method to detect the recall rate of low resolution license plate recognition. The results are shown in Fig. 4.



**Fig. 4.** Recall rate of low resolution license plate recognition under different methods

Analysis of Fig. 4 shows that the recall rate of low resolution license plate recognition is different under different methods. When the number of vehicles to be identified is 40, the low resolution license plate recognition recall rate of literature [7] method can reach 73%, the low resolution license plate recognition recall rate of literature [8] method can reach 62.5%, the low resolution license plate recognition recall rate of literature [9] method can reach 49%, and the low resolution license plate recognition recall rate of this method can reach 87.5%. When the number of vehicles to be identified is 60, the recall rate of low resolution license plate recognition of literature [7] method can reach 77%, the recall rate of low resolution license plate recognition of literature [8] method can reach 74%, the recall rate of low resolution license plate recognition of literature [9] method can reach 60%, and the recall rate of low resolution license plate recognition of this method can reach 90%. This method has higher recall rate of low resolution license plate recognition, which shows that the license plate recognition effect of this method is better.

## 5 Conclusion

This paper proposes a low resolution license plate recognition method based on intelligent data processing and prediction algorithm. The low resolution digital image of license plate is defined by the principle of image registration; The Gaussian pyramid and Gaussian difference pyramid models of low resolution license plate are constructed, and the false matching of low resolution license plate is eliminated by RANSAC prediction algorithm to realize low resolution license plate recognition. The results are as follows:

- (1) When the number of iterations is 20, the low resolution license plate recognition accuracy of this method is 90%. The low resolution license plate recognition accuracy of this method is much higher than other methods.
- (2) When the iteration times are 30, the time of low resolution license plate recognition in this paper is 89s. The low resolution license plate recognition time of this method is far less than the other two methods.

- (3) When the number of vehicles to be recognized is 60, the recall rate of low resolution license plate recognition can reach 90%. This method has higher recall rate of low resolution license plate recognition, which shows that the license plate recognition effect of this method is better.

**Fund Projects.** 1. Project name: incomplete license plate recognition study based on generative adversarial network, project number: KJQN201905503  
 2. Project name: The work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission, project number: KJZD-K201801901  
 3. Project name: research and design of intelligent manufacturing cloud platform system based on Internet of things, project number: KJQN201801902.

## References

1. Zhu, Q., Liu, S., Guo, W.: Research on license plate detection based on FASTERR-CNN. *Auto Ind. Res.* **296**(01), 59–62 (2019)
2. Tang, Y.: License plate recognition algorithm based on fusion of gradient operator and mathematical morphology. *Electron. Technol. Softw. Eng.* **187**(17), 130–131 (2020)
3. Zheng, G., Wu, H.: Method of license plate location based on MSER feature and edge projection. *Comput. Eng. Des.* **40**(01), 249–252 (2019)
4. Liu, N., Zhang, J.: Chinese license plate recognition system based on Haar features. *J. Jimei Univ. (Nat. Sci.)* **24**(2), 139–144 (2019)
5. Wang, Z., Ma, X., Huang, W.: Vehicle license plate recognition based on wavelet transform and vertical edge matching. *Int. J. Pattern Recognit. Artif. Intell.* **34**(06), 1134–1142 (2020)
6. Sathya, K.B., Vasuhi, S., Vaidehi, V.: Perspective vehicle license plate transformation using deep neural network on genesis of CPNet. *Procedia Comput. Sci.* **171**(1), 1858–1867 (2020)
7. Onim, M., Akash, M.I., Haque, M., et al.: Traffic surveillance using vehicle license plate detection and recognition in Bangladesh (2020)
8. Swastika, W., Sakti, E., Subianto, M.: Vehicle images reconstruction using SRCNN for improving the recognition accuracy of vehicle license plate number. *Jurnal Teknologi dan Sistem Komputer* **8**(4), 304–310 (2020)
9. Islam, R., Islam, M.R., Talukder, K.H.: An efficient method for extraction and recognition of Bangla characters from vehicle license plates. *Multimed. Tools Appl.* **79**(27), 20107–20132 (2020)
10. Al-Shemarry, M.S., Li, Y., Abdulla, S.: An efficient texture descriptor for the detection of license plates from vehicle images in difficult conditions. *IEEE Trans. Intell. Transp. Syst.* **21**(2), 553–564 (2020)
11. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–17 (2019)
12. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. *Multimed. Tools Appl.* **76**(4), 5787–5802 (2014). <https://doi.org/10.1007/s11042-014-2408-1>
13. Liu, S., Bai, W., Liu, G., et al.: Parallel fractal compression method for big video data. *Complexity* **20**(18), 1–16 (2018)
14. Luo, L.Z., Fei W.U., Cao, K., et al.: Application of image super-resolution in fuzzy license plate recognition system. *Softw. Guide* **18**(05), 177–180+186 (2019)

15. Hu, X., Li, X.S., Li, Y., et al.: Research and implementation of blind restoration algorithm for moving fuzzy license plate image based on frequency-domain characteristics. *Int. J. Pattern Recogn. Artif. Intell.* **25**(6), 36–40 (2021)
16. Chowdhury, M., Dhar, P.: License plate detection based on fuzzy rule. *Int. J. Adv. Sci. Technol.* **29**(9), 8583–8590 (2020)



# Design and Application of Security Monitoring System for Perception Terminal of Power Internet of Things

Hong Xu<sup>1</sup>(✉), Xin Sun<sup>2</sup>, Jie-yao Ying<sup>3</sup>, and Qing-li Niu<sup>4</sup>

<sup>1</sup> School of Software, Zhengzhou University, Zhengzhou 314500, China  
xuhhhh06@outlook.com

<sup>2</sup> School of Electrical Engineering, Zhejiang University,  
Hangzhou 310000, China

<sup>3</sup> School of Software, Zhejiang University, Hangzhou 310000, China

<sup>4</sup> College of Information Engineering, Zhengzhou University of Science  
and Technology, Zhengzhou 450064, China

**Abstract.** The information acquired by the traditional power Internet of Things sensing terminal monitoring system is not comprehensive, which leads to the long time needed to monitor the terminal security vulnerabilities. In order to solve this problem, this paper designs a new security monitoring system of power IOT sensing terminal. On the basis of the traditional system hardware design, the overall architecture of the system is designed by fully considering the security function of the power IOT sensing terminal. Then, the data of sensing terminal is collected at the monitored object end, and the data display module is designed. Then, the specific monitoring module is designed from three aspects of monitoring process, power Internet of things sensing terminal application, client and server, so as to realize the security monitoring of power Internet of things sensing terminal. In the experimental part, we design the vulnerability of the power IOT sensing terminal and the possible attack actions under the vulnerability. The results show that compared with the traditional system, the system in this paper needs less time to monitor the vulnerability of the power IOT sensing terminal.

**Keywords:** Power Internet of things · Perception terminal · Security monitoring system · Data acquisition · Security vulnerability

## 1 Introduction

In order to fully and effectively implement the planning, design and management of power network, and realize the unified utilization of network information, the State Grid Corporation of China proposed the deployment arrangement for the comprehensive construction of ubiquitous power Internet of things [1]. The implementation of power Internet of things also puts forward new requirements for comprehensive perception of perception layer devices. Therefore, relevant experts design ubiquitous perception terminal installed on the user side on the basis of the original. However, the network communication between the Internet is usually protected by firewall and

intrusion detection system, and the monitoring inside the network is often forgotten, which is also the weak point of prevention [2]. Therefore, the current design of the power Internet of things perception terminal, there are non external way to infect the virus, from the internal deployment of the virus or Trojan horse to destroy the normal network communication, hacker attacks and other security risks.

In foreign countries, for IOT sensing terminal monitoring, equipment monitoring and fault diagnosis consultation and technology promotion are generally carried out through the network, and some information exchange formats and standards are formulated. Due to the rapid development of computer technology and communication technology, in recent years, China has begun to actively carry out this research [3]. Reference [4] divides the monitoring of Internet of things into two layers. The upper layer uses standard Ethernet, the lower layer uses s-485 protocol bus technology, plus servers and monitoring workstations to form a LAN monitoring system suitable for industrial field. Reference [5] a WWW browser is installed on the user terminal to obtain the online technical support and data exchange from the TSB of the remote service department through HTTP to complete the security monitoring of the Internet of things perception terminal. However, the information obtained by the traditional power Internet of things sensing terminal monitoring system is not comprehensive, which leads to a long time to monitor the terminal security vulnerabilities.

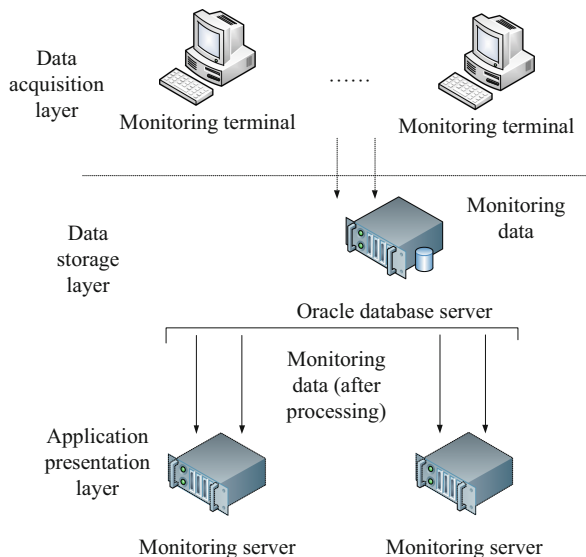
In view of the shortcomings of the traditional system, this study designed a new power Internet of Things sensing terminal security monitoring system. The design idea of the system is as follows:

- (1) Considering the existing problems in the security of the sensing terminal of the electric Internet of Things and the status quo of the hardware design of the monitoring system, the hardware and software parts of the system are designed on the basis of clarifying the security functions of the sensing terminal.
- (2) Design the data display module and monitoring module, and complete the security monitoring of the sensing terminal of the power Internet of Things from the perspectives of monitoring process, sensing terminal application, client and server.
- (3) The design experiment proves that the system in this paper needs less time to perceive vulnerabilities.

## **2 Design of Security Monitoring System for Perception Terminal of Power Internet of Things**

### **2.1 System Architecture Design**

The security monitoring system of the power Internet of things perception terminal designed in this study mainly includes three modules: monitoring terminal software, monitoring data storage software and integrated display and control software. The functions of information monitoring, data storage and comprehensive display are processed respectively. The structure is shown in Fig. 1.



**Fig. 1.** Design of network monitoring system

As shown in Fig. 1, in the architecture of the monitoring system, the data acquisition layer monitoring terminal software mainly completes the monitoring data acquisition of the terminal and the data upload function after the data acquisition; the data storage layer completes the data storage, and the user can save the collected historical monitoring data and complete the data processing on this basis, which serves as the interface between the data acquisition layer and the application presentation layer. In addition to the above two functions, the data storage layer completes the preliminary data analysis and processing in the data acquisition layer, and completes the secondary data analysis and processing through stored procedure programming, which greatly improves the data processing speed of the application presentation layer. Stored procedure programming is a database programming technology, It has the advantages of simplicity, security and high performance; the application presentation layer mainly completes the display of the data collected by the data collection layer, as well as the functions of data analysis, early warning and statistical report.

## 2.2 Collect the Data of the Sensing Terminal

The system achieves the acquisition of monitoring information by installing a terminal in the monitored object end, and then the monitoring terminal uploads the collected information to the monitoring server, so as to achieve the acquisition of information. Therefore, it is necessary to install the client in the monitoring terminal to fully control the monitoring terminal, and it will occupy a certain amount of monitoring terminal resources. The acquisition process is shown in Fig. 2.

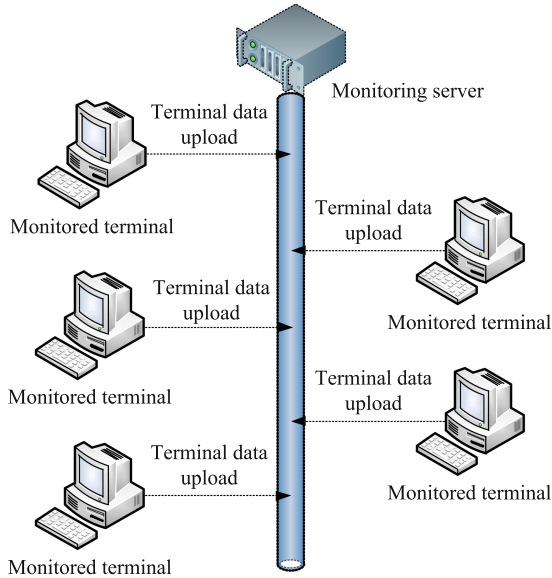


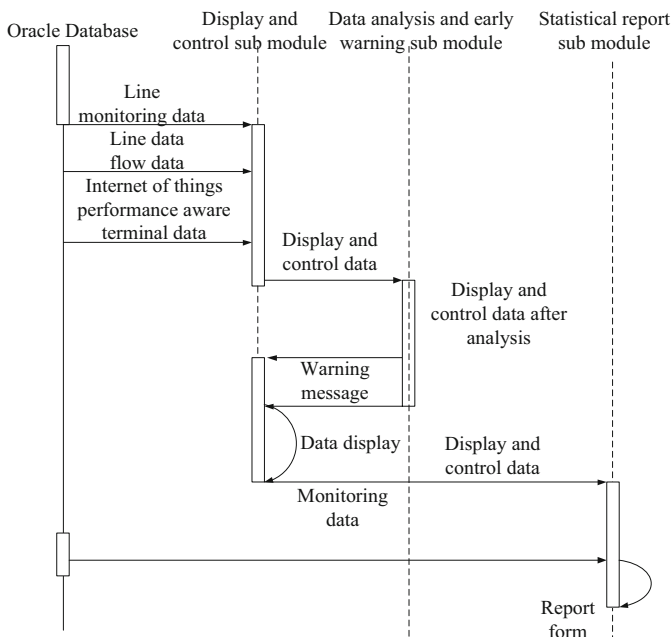
Fig. 2. Sensing terminal data acquisition

Because the monitoring terminal is installed with a client, when a monitoring terminal crashes or the network delays, it only leads to the disconnection of a single node and does not affect the monitoring server [6]. In addition, when the monitoring terminal changes (increases and decreases), additional operation of installing the monitoring terminal is needed. Secondly, the monitoring terminal needs to upload data to the monitoring server. When the monitoring server is installed with firewall, the access of the monitoring terminal will be limited.

### 2.3 Data Display Module

The data display module mainly completes the display of the data collected by the data acquisition layer, including three functions of monitoring data acquisition, monitoring data display and control monitoring terminal. Its structure is shown in Fig. 3.





**Fig. 3.** Sequence diagram of data display software

In the software timing diagram of data display module shown in Fig. 3, the monitoring data acquisition is the interaction between display and control sub module through data interaction with Oracle database. The data including line monitoring data, line data flow data and equipment performance data, and data that have been analyzed and processed from the database storage process are obtained.

Monitoring data display is that after the display and control sub module obtains the monitoring data from Oracle database, it respectively realizes the line monitoring display, line data flow display and equipment performance display. The display and control sub module submits the obtained data to the data analysis and early warning sub module for data analysis, and carries out data early warning display according to the feedback results of the data analysis and early warning sub module.

After getting the monitoring data from Oracle database, the display and control sub module realizes the line monitoring display, line data flow display and equipment performance display respectively. The display and control sub module submits the obtained data to the data analysis and early warning sub module for data analysis, and carries out the data early warning display according to the feedback results of the data analysis and early warning sub module.

The control and monitoring terminal mainly includes monitoring terminal configuration modification and monitoring terminal stop, restart and resume control operations. The control and monitoring terminal completes the control of the monitoring terminal through the terminal link security confirmation.

## 2.4 Design of Monitoring Module of Monitoring System

### System Monitoring Process

Data acquisition layer monitoring terminal software mainly completes the monitoring data acquisition and data upload function after data acquisition [7]. The data acquisition layer includes network lines, line data flow and network equipment related performance information. According to the functions to be realized, it is divided into five sub modules: system tray module, line monitoring sub module, line data flow monitoring sub module, equipment performance monitoring sub module and monitoring data storage module. Based on these five sub modules, the software module flow of terminal security monitoring system is designed, as shown in Fig. 4.

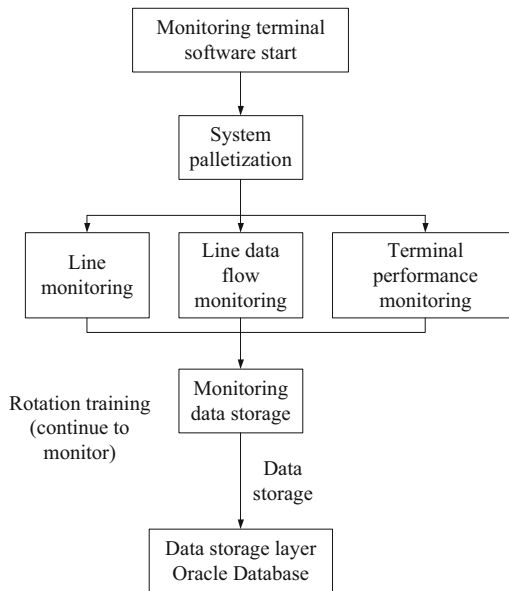


Fig. 4. Software module flow of terminal security monitoring system

As shown in Fig. 4, the software module flow of terminal safety monitoring system. After the system starts the system software, the software tray will be displayed in the system tray program at the bottom right of the system, which is used to identify the current monitoring terminal software in operation. Once the monitoring terminal program is abnormal, the system tray will disappear.

After the software tray of monitoring system is started, three monitoring threads will be started respectively: line monitoring thread, line data flow monitoring thread and device performance monitoring thread. In order to understand, this study is called sub module.

The software of monitoring system adopts three threads to realize the realization of the sub module of line monitoring, the sub module of line data flow monitoring and the performance monitoring module of equipment. The three threads perform the functions of these three sub modules respectively [8].

The three monitoring threads will simultaneously upload the acquired line monitoring data, line data flow monitoring data and device performance monitoring data to the data storage layer Oracle database through the monitoring data storage module.

### **Terminal Monitoring**

Monitoring the power Internet of things perception terminal can be divided into two parts: the server and the client, respectively monitoring the power Internet of things perception terminal. The server is composed of security monitoring module, information display module, instant messaging module, integrated tool module and central database.

Security monitoring module includes: LAN scanning processing module, port summary processing module, filter analysis processing module, log processing module, firewall processing module, etc. The monitoring server dynamically analyzes all the incoming and outgoing IP addresses and ports in the LAN. According to the abnormal analysis of the ports and incoming and outgoing packets, the illegal operation can be judged. Summarize the ports by IP address to view the illegally used network processes. The abnormal IP address can be analyzed separately to determine the main object of illegal operation. Firewall can be used to block the port and address.

Information display module includes: screen capture processing module, client locking module, communication processing module, client management module, etc. It mainly monitors the client on a regular basis, and remotely captures the abnormal traffic. The client can directly view the screen. If illegal use is found, the message module will give a warning. If you don't pay attention to it, you can lock the client or remotely close the client.

Instant messaging module. It is mainly aimed at content and message capture in abnormal time, such as communication software.

The integration tool module includes: routing test and connectivity test. Using these tools to test the connectivity of local LAN and the function of remote network routing, the administrator can quickly find the cause of network failure.

Central database. It mainly stores and manages all kinds of data, and is the storage center of system monitoring data.

The client is composed of the message part and the request processing module sent by the local receiving server. Through the interaction between these menus, the integrity of the whole system can be guaranteed [9]. The client mainly includes the following parts:

Information capture part: mainly responsible for collecting the key system information (CPU, memory, process list, etc.) on the host.

Communication module: the main function is to receive instructions or messages sent by the server, make corresponding responses, such as lock/unlock, shutdown and other instructions, and realize encrypted communication with the server.

Hiding module: it is mainly responsible for the automatic loading of the client and the automatic hiding of the corresponding process.

## Application Monitoring

When a computer starts an application, the operating system creates a process for the program. When the process is terminated, the application ends. Each process in Windows has its own address space, which contains the code and data of all executable modules or dynamically linked library modules. Only when the windows operating system creates a process for it and allocates the necessary resources, the program can run. For Windows environment, whether you double-click the application with the mouse or enter the command in the command line mode, the corresponding application creation is completed by the operating system process calling the windows API function. In this way, you can filter the corresponding application by monitoring the CreateProcess function. DLL has always been the foundation of Windows operating system. Functions in the windows API are included in the DLL. The three most important DLLs are Kernel32. DLL, User32. DLL and GDI 32. DI. The use of DLL can save memory space very effectively. If multiple applications use a DLL together, the pages of the DLL will only be put into memory once, and all applications can share pages. And the resources in the DLL can be shared by all these applications. Because of these characteristics of DLL, we can monitor some API functions, and only DLL can inject into the address space of all processes.

HOOK API is an effective and simple way to monitor the application process. Hook is a mechanism for Windows to process messages. An application program can monitor some messages in a specified window by setting its subprogram, and the monitored window can be created by other processes [10].

Microsoft does not provide the relevant information of hook API, but from the characteristics of DLL and the functions provided by HOOK, it is not particularly difficult to implement hook API. In fact, the development of hook API technology has been very mature.

However, it should be noted that due to the problem of character set, there are two versions of a large part of Windows API functions: the version suitable for UNICODE and the version suitable for DBCS, which correspond to CreateProcessW of wide character set and CreateProcessA of double byte character set.

Since the operating system makes a large number of calls to system API that need to be monitored, in order to improve the response speed of the operating system, two methods are adopted to improve the efficiency of the operating system: the speed of accessing security policy files and the speed of matching rules.

During the policy check, if the normal file reading and writing method is adopted, the file needs to be opened and closed frequently, and the file is frequently read and written, which will reduce the efficiency of the operating system. Therefore, this system adopts the method of memory file mapping. Even if the application program accesses the file on the disk through the memory pointer, the process is to access the memory of the loaded file. When using memory-mapped files for I/O processing, the operating system performs data transfer pages. As for all internal pages, the virtual memory manager is responsible for management, because the virtual memory manager is a unified method It handles all disk I/O, so this optimization makes it capable of processing memory operations at the fastest speed.

Specifies the matching speed. The security policy rules are not arranged in disorder, but are classified according to the disk drive where the monitored object is located, and

arranged in alphabetical order. When the security policy is checked, what is known is the name of the monitored object, that is, the name of the file or folder, and the disk drive where the monitored object is located can be obtained according to the file name. The disk drive can be used to directly locate the scope of the strategy applicable to the monitored object according to the corresponding location index of the file, thereby greatly reducing the number of rule matching checks that need to be performed.

### 3 System Test

A comparative test is adopted to verify the effectiveness of the security monitoring system for the sensing terminal of the power Internet of Things designed in this paper. The perception terminal of the power Internet of Things in a certain area is taken as the experimental object. In order to enhance the contrast between the experimental results, the two traditional security monitoring systems were compared, and the time required to monitor the vulnerability of the sensing terminal of the power Internet of Things was taken as an indicator to verify the effectiveness of the different systems.

#### 3.1 Experiment Preparation

Load the Rapid SCADA software, which can provide information to the power dispatcher, and run it on the Kali virtual machine on Windows Server 2020 system and several station-controlled virtual monitoring layers to complete the construction of the test platform.

On the perception terminal of the electric Internet of Things, the following three vulnerabilities are set:

1. In the case of unauthorized access, files that cannot be accessed are accessed, and file transmission occurs. The rhost vulnerability of the protocol, and the vulnerability is recorded as  $a_1$ ;
2. Allow illegal users to send forged Modbus data packets, and the Modbus protocol authorization vulnerability appears, and the vulnerability is recorded as  $a_2$ ;
3. Execution without permission Power Internet of Things scheduling command, a WinCC software vulnerability appears, and the vulnerability is recorded as  $a_3$ .

According to the three vulnerabilities set above, it is assumed that the power dispatching network may be subjected to the following attack actions:

1. Use vulnerability  $a_1$  to obtain user rights and record it as a  $b_1$  attack action;
2. Use vulnerability  $a_2$  to deceive the network protocol and record it as a  $b_2$  attack Action;
3. Use vulnerability  $a_3$  to tamper with terminal data and record it as  $b_3$  attack action.

On this basis, the perception terminal of the electric Internet of Things is set to three operating states, as follows:

1. Normal operating state;
2. Obtaining user authority state;

3. Manipulating the state of power dispatching network nodes.

In the above three states, the perception terminal information of the power Internet of Things was subjected to three attacks, and then the time required by the three system monitoring vulnerabilities was tested.

3.2 Experimental Result

The First Set of Experimental Results

The power Internet of Things sensing terminal is set to normal operation state, and then the security state of the power Internet of Things sensing terminal is monitored under different attack actions respectively, and the time required for the three system monitoring vulnerabilities is recorded. The experimental results are shown in Fig. 5.

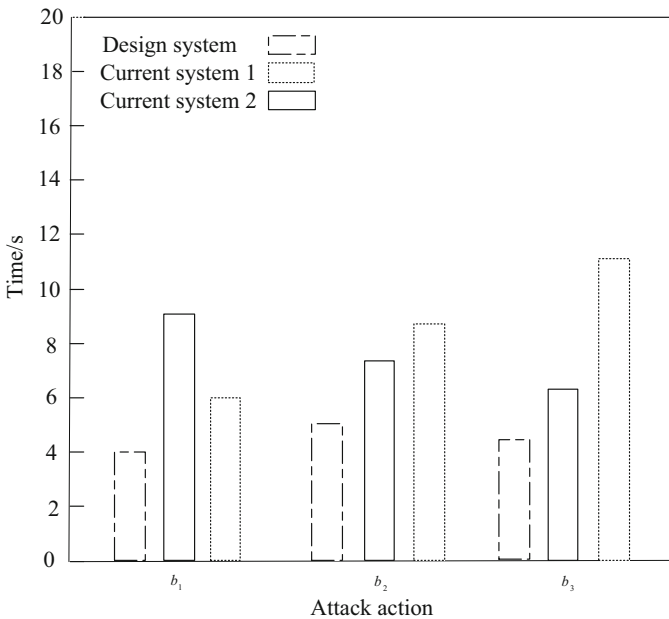
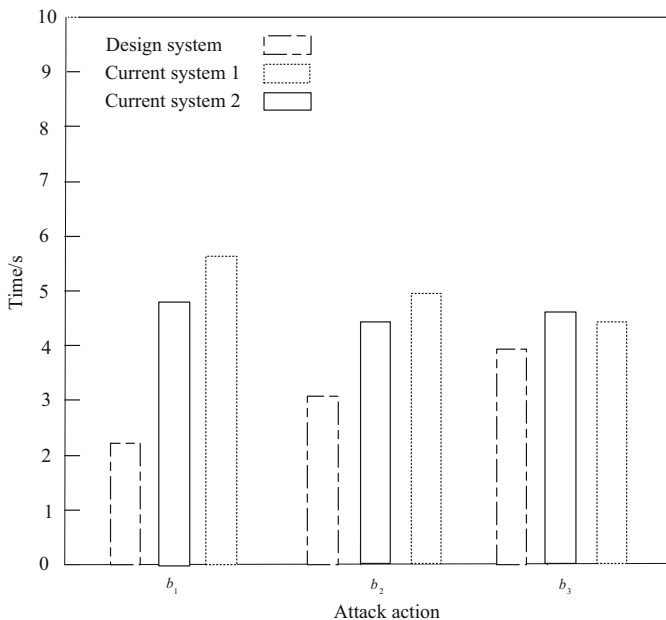


Fig. 5. The time required for different systems to monitor vulnerabilities under normal operating conditions

As can be seen from Fig. 5, under the  $b_1$  attack action, it takes more time for system 1 to monitor the vulnerability than System 2. Under attack actions  $b_2$  and  $b_3$ , it takes less time for system 1 to monitor the vulnerability than System 2. However, the monitoring time of these two systems is always longer than that of the system in this paper, which proves that the system in this paper is obviously superior to System 1 and System 2.

### The Second Set of Experimental Results

The power Internet of Things sensing terminal is set to obtain the user's authority status, and then the security state of the power Internet of Things sensing terminal under different attack actions is monitored respectively, and the time required for the three system monitoring vulnerabilities is recorded. The experimental results are shown in Fig. 6.

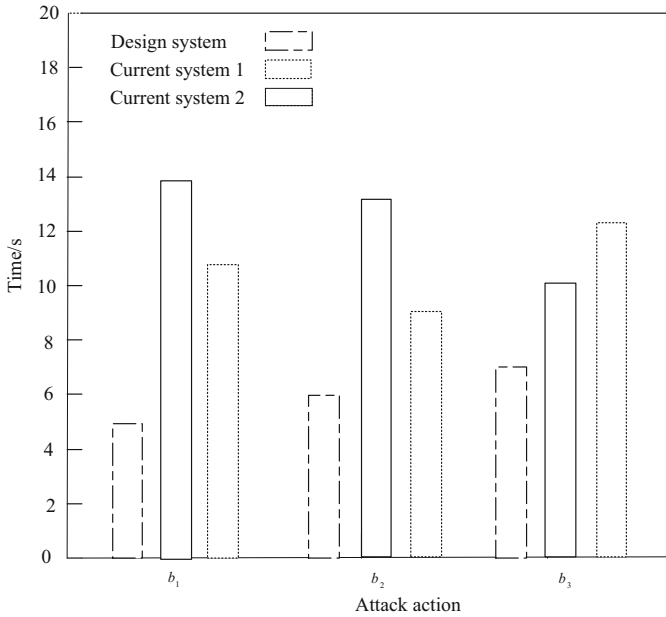


**Fig. 6.** Obtain the time required for monitoring vulnerabilities of different systems under the status of user permissions

As can be seen from Fig. 6, when the perception terminal of the power Internet of Things is in the state of obtaining user rights, the time required for monitoring vulnerabilities of the three systems is significantly reduced compared with that in the normal operation state. The average monitoring time of system 1 is 5 s, which is the longest. The average monitoring time of system 2 is 4.7 s, which is still relatively long. The average monitoring time of the system in this paper is 3.1 s, which is obviously more effective than that of System 1 and System 2.

### The Third Set of Experimental Results

The power Internet of Things sensing terminal is set to manipulate the state of power dispatching network nodes, and then the security state of the power Internet of Things sensing terminal is respectively monitored under different attack actions, and the time required for the three system monitoring vulnerabilities is recorded. The experimental results are shown in Fig. 7.



**Fig. 7.** The time required to restore stable operation of the network under the control of the state of the power dispatch network node

As can be seen from Fig. 7, when the sensing terminal of the power Internet of Things is in the state of manipulating the power Internet of Things nodes, the time required for monitoring vulnerabilities of the three systems is significantly increased compared with that in the normal operation state, indicating that the process of monitoring vulnerabilities is the most difficult at this time. In this state, the average monitoring time of system 2 is longer, up to 13 s. System 1 has the longest average monitoring time, up to 14 s. The maximum monitoring time of the system in this paper is 7.2 s, which proves that it is better than system 1 and system 2.

Causes of these results is that the system takes into account the problems existing in current electric power iot terminal security perception as well as the monitoring and control system hardware design present situation, in a clear perception terminal was designed on the basis of the safety function of system software and hardware parts, and the monitoring process, perception terminal application, the client and the server a few point of view, To complete the comprehensive and comprehensive security monitoring of the perception terminal of the electric Internet of Things, so as to reduce the time needed for its perception vulnerability.



## 4 Conclusion

In this study, a security monitoring system for the sensing terminal of the power Internet of Things is designed. Considering the existing problems in the security of the sensing terminal of the power Internet of Things and the status quo of the hardware design of the monitoring system, the hardware and software parts of the system are designed on the basis of clarifying the security functions of the sensing terminal. Then, the data display module and monitoring module are designed to complete the security monitoring of the sensing terminal of the power Internet of Things from the perspectives of monitoring process, sensing terminal application, client and server. This study also proved that the system needed less time to perceive vulnerabilities through comparative experimental results, thus proving the effectiveness of the system.

But the design of the security monitoring system, there are still some deficiencies. In the future research, it is also necessary to further study the weak links of the perception terminals of the electric Internet of Things and increase the practical functions of the monitoring system.

## References

1. Lin, H., Luo, W.: Internet of things terminal monitoring based on network data analysis. *Electron. Des. Eng.* **28**(18), 101–105 (2020)
2. Huang, S., Zhang, X.: Design and implementation of monitoring system based on network terminal. *Digital Technol. Appl.* **37**(7), 94–95 (2019)
3. Lu, Q., Cui, W.: Security monitoring analysis technology of terminal layer in ubiquitous power IoT. *Inf. Technol.* (2), 121–125+134 (2020)
4. Xiao, A., Zhang, W., Zhao, D., et al.: Construction method of smart grid monitoring system based on IoTs. *Inf. Technol.* (12), 86–90+95 (2020)
5. Chen, J., Zhou, Z., Feng, W., et al.: Design of power monitoring system based on wireless network. *Tech. Autom. Appl.* **39**(4), 168–171 (2020)
6. Du, L.: Design of wireless temperature monitoring system for power cable. *J. Beijing Polytech. Coll.* **19**(1), 23–27 (2020)
7. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. *Front. Genet.* **10**, 120 (2019)
8. Lei, Y.: Real-time monitoring system with wide spectrum and multiple parameters based on Internet of Things. *Comput. Netw.* **46**(6), 62–65 (2020)
9. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
10. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. *Entropy* **21**(9), 902 (2019)



# Reliability Evaluation Model of Online Teaching Quality Based on Big Data Technology

Feng Wang<sup>(✉)</sup>

Jiujiang University, Jiujiang 332005, China  
wangfengl7172021@163.com

**Abstract.** In order to ensure the education ability of online teaching course and solve the disadvantages of traditional teaching methods in location, a reliability evaluation model of online teaching quality based on big data technology is designed. By classifying the online teaching quality assessment method, comparing the practical application ability of each evaluation method, we determine a most reliable evaluation method, and complete the construction of online teaching quality evaluation index system. On this basis, a perfect online teaching system is built. According to the application requirements of evaluation framework and evaluation strategy, the necessary reliability evaluation objectives are defined, and the online teaching quality reliability evaluation model based on big data technology can be successfully applied. Comparing the model with the traditional structural equation model, it can be seen that big data technology can strengthen the reliability evaluation of online teaching quality by network hosts, thus breaking the inhibition effect of location restrictions on online curriculum teaching ability.

**Keywords:** Big data technology · Online teaching · Reliability assessment

## 1 Introduction

Big data, as an IT industry term, refers to the data set that cannot be captured, managed and processed by conventional software tools within a certain period of time. It is a massive, high growth rate and diversified information asset that requires new processing mode to have stronger decision-making power, insight and process optimization ability. The strategic significance of big data technology is not to master huge data information, but to professionally process these meaningful data. In other words, if big data is compared to an industry, the key to making profits in this industry lies in improving the “processing ability” of data and realizing the “value-added” of data through “processing” [1–3]. Big data needs special technologies to effectively handle a large amount of data over time. Technologies suitable for big data include massive parallel processing (MPP) database, data mining, distributed file system, distributed database, cloud computing platform, Internet and scalable storage system. Big data includes structured, semi-structured and unstructured data, and unstructured data is increasingly becoming the main part of data.

Compared with the traditional offline teaching, the online teaching mode can meet more students' classroom learning at one time. The teachers' original need for on-site teaching has also become on-site teaching. The same number of students, who may need to teach in batches for various reasons, has also become one-time teaching, which greatly improves the time efficiency. But at the same time, under the bright appearance, it also has its own shortcomings. In the common practical application, because students and teachers are separated from each other, it is difficult to communicate directly across the screen, resulting in no targeted guidance, or in the face of rich resources do not know how to screen and use, a separate learning environment leads to the neglect of others, at the same time, they can not integrate into the learning society, and then reduce the learning efficiency. To sum up, online education has its own problems. How to evaluate the performance of online teaching is a problem worthy of study.

In this paper, a higher vocational college as an example, the university will set up a number of courses as its own online education pilot courses, organize students to study at fixed time, although the convenience of MOOCS is not enough, but this is a good solution to the problem of low efficiency of individual learning. But at the same time, whether online education can change the dilemma of our traditional education, whether it is the future development and reform of education, how many advantages this new online education mode has, and whether this mode can be accepted by teachers and students, this paper will use the analytic hierarchy process to establish a judgment matrix, set index weight and various scores, and draw a conclusion. Therefore, it can be used as a reference for the reform and innovation of teaching mode in Colleges and universities.

## 2 The Construction of Online Teaching Quality Evaluation Index System

### 2.1 Various Methods to Evaluate the Quality of Online Teaching

Effective evaluation of teachers' teaching quality can enable teachers to give full play to their awareness of educational innovation in the classroom and achieve the effect of different educational concepts. Evaluation of teaching quality can identify efficient teaching methods. Education reform is to bring education to a higher and better platform. Education reform has  $B$  clear direction through the invisible "eyes" of teaching quality evaluation. Let  $a$  represent the AHP evaluation coefficient of online teaching, and  $X$  represent the fuzzy evaluation coefficient of online teaching:

$$C = \left( \frac{1}{B\sqrt{y} \cdot X\sqrt{p}} \right)^{\frac{3}{n}} \quad (1)$$

Among them,  $y$  and  $p$  represent two different online teaching quality evaluation coefficients. Teaching quality evaluation refers to taking scientific evaluation methods according to certain evaluation standards, following strict evaluation procedures, considering and evaluating the work results and behaviors of individuals or organizations, and obtaining quantitative evaluation data and results, so as to guide the work

tasks in the next period. In the specific performance evaluation, we can focus on the following issues: ① whether we understand the evaluation object; ② whether we understand the purpose of the evaluation; ③ according to what standards to carry out the evaluation; ④ whether we have sufficient relevant information consistent with the evaluation purpose; ⑤ whether the evaluation results are objective and accurate; and ⑥ whether the evaluation results can promote the evaluation object to make clear the direction of future efforts.

Big data theory holds that the management of online teaching quality evaluation is not only the summary and test of the previous management work, but also the premise of problem discovery and solution. This is to improve the management efficiency and enhance the production efficiency of the organization. It is a general evaluation of the comprehensive operation effect of the organization's management by using relevant indicators and systems. Generally speaking, the basic content of online teaching management is roughly divided into performance planning, performance communication, data analysis and performance evaluation. Teaching quality evaluation should be a comprehensive index system, a complete system and a dynamic process. At present, the construction of online teaching quality evaluation system needs to follow the basic themes of efficiency, effect, economy and fairness. For performance management, performance evaluation is not only a basic project, but also a constraint mechanism to promote management. It can not only adjust the orientation of new goals, but also highlight the value orientation of performance management.

## 2.2 Comparison of Evaluation Methods

Teaching evaluation is an effective way to understand and analyze the current situation of teacher education. Only after the evaluation of teaching quality can teachers show their advantages and disadvantages. Only by the evaluation of teaching quality can teachers have the reference objects for the development of educational and teaching, which can promote the peer to discuss, learn and promote each other. And create a kind of atmosphere between teachers to pursue each other. Through the evaluation of teaching, teachers can not only understand the advantages and disadvantages of their own education, but also enhance the motivation of internal communication between teachers to listen to and evaluate the lessons, and promote teaching research. Set  $\chi$  to represent the weighted coefficient of online teaching,  $f$  represents the characteristic value of online teaching contribution proportion, and establish a formula (1). The comparison results of teaching quality of different evaluation methods can be expressed as follows:

$$L = \frac{\sqrt{(\chi - 1)|r \cdot w|}}{C \cdot f^2} \quad (2)$$

where  $r$  is the subjective weighted value and  $w$  is the passive weighted value. Different types of colleges and universities focus on the evaluation of teachers' teaching quality, so the evaluation index will be different. Therefore, the construction of a broader teaching quality evaluation model mainly depends on the selection of index variables, which are usually scientific, comprehensive and testable. Usually, the index variables

are mostly selected from teaching methods, teaching content, teaching means and teaching effect. It is easy to repeatedly select and use some indicators from several evaluation angles, which often leads to different evaluation results because of different evaluation angles.

### 2.3 Determination of Online Teaching Reliability Evaluation Method

Big data technology model is a statistical analysis method, which uses covariance between variables to study the causal relationship of variables. The structural equation model can not only express the linear relationship of measurable variables, but also the potential variables that can not be observed; The direct and indirect relationship between variables can be expressed, so the structural equation model is more feasible than the traditional evaluation method. However, structural equation model is also mentioned as a verification factor analysis method, which usually needs the initial model with theory or experience as support [4–6]. When reading a large number of literature, it is found that the integration of the original index system by factor analysis method not only overcomes the subjectivity of the researchers in selecting the index system of the evaluation object, but also lays the theoretical foundation for the structural equation model to evaluate the teaching quality of teachers in class.

Teaching level and ability are the foothold of teachers. One of the most important topics in teacher education is how to effectively improve teachers' teaching level and ability. Obviously, the good evaluation of classroom teaching quality has become a pair of eyes, which can provide scientific understanding for most teachers' teaching situation. Let teachers understand the shortcomings of teaching and the direction of future efforts, and guide the improvement of teachers' professional quality. Teaching evaluation is not only used to evaluate the work of teachers, but also to make the quality of teachers more detailed, and to distinguish the professional ability, attitude and teaching quality of a teacher through objective and fair quality assessment. Through processing a large number of raw data generated by students online evaluation, the teachers' hard work is displayed scientifically and fairly. At the same time, it also provides system guarantee for more students to meet a good teacher who is conscientiously responsible. Set  $v_1$  and  $v_2$  to represent two different original online teaching data, and establish a formula (2), which can express the reliability evaluation results of online teaching as follows:

$$T = \frac{|L|}{\sum_{i_0} g(v_1 + v_2)} \quad (3)$$

Among them,  $i_0$  represents the minimum missing value of reliability evaluation information, and  $g$  represents the contribution coefficient of reliability evaluation. With the support of big data technology, online teaching quality evaluation refers to making an objective, fair and accurate comprehensive evaluation on the operation benefit and the performance of the operators during a certain operation period of the project by using the principles of mathematical statistics, operational research and specific index system, according to the unified standards and certain procedures, through quantitative

and qualitative comparative analysis. The results of performance evaluation can directly affect the vital interests of many employees, such as salary adjustment, bonus payment and job promotion.

### **3 Reliability Evaluation Model Based on Big Data Technology**

#### **3.1 Construction of Online Teaching System**

Online teaching is not only a safe and reliable online teaching platform, but also the school system uses reasonable programming and simple technology to realize. The purpose is to create a platform for teachers and students to learn and communicate easily and happily. Teachers can quickly operate the platform, spend more time on the design of teaching content and the development of teaching activities. Students can watch the live teaching course through the student client. The teaching process is no longer just video, and the interactive communication between the live broadcast and the teaching teachers and other students is used to solve the difficult problems encountered in the learning process by online questions and answers. System management personnel can easily check the running state of the system, reduce the difficulty of management technology and management cost, improve the efficiency of management, and better promote the work of online education in schools.

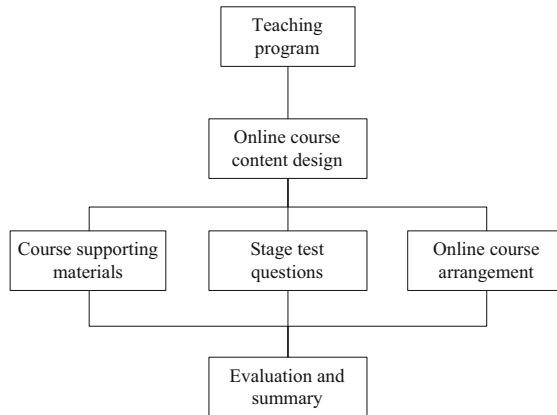
In summary, the main means of big data system establishment are mature computer system and advanced information communication technology, which can make students, teachers and managers meet the requirements of high efficiency, high quality and good operation effect interactive teaching platform. The construction characteristics of the system are safety, reliability, convenient operation, strong practicability and strong maintainability.

The following principles must be followed in the establishment of a perfect online teaching quality reliability evaluation system:

- (1) Reliability principle. In any system, the first thing to ensure is the reliability of the system. The system has a certain fault tolerance rate, and there must be a backup scheme for the key data of the system platform, so as to facilitate the fault recovery in an extraordinary period. Once there is an accident, it can ensure the data security to the maximum extent, and have the means of fast and safe recovery of data, so as to reduce or even avoid the loss caused by the loss of key data.
- (2) The principle of concurrent processing ability. When the system is running, it must have the ability to bear a large number of visits and concurrent requests. When selecting the big data server used by the system, it is necessary to pay attention to whether the performance of the selected server can meet the processing of a large number of concurrent requests and a good queuing mechanism, so as to prevent the system from no response due to excessive visits during the course.
- (3) The principle of advanced nature. The use of hardware and architecture system of the system must be in line with the advanced level of the industry, to ensure that the platform system has a long vitality, at the same time has a certain advance and development potential, and can be used for a long time, in line with the current

and future needs. The advanced nature of the system is mainly reflected in advanced concepts, advanced technology, advanced hardware and advanced software architecture.

- (4) The principle of easy management. Generally speaking, after the successful operation of the system, the follow-up emphasis should be placed on maintenance and management. This requires that in the initial design of the big data platform, the need for future system maintenance and management should be taken into account, so as to reduce the unnecessary capital and personnel investment in the later maintenance, so as to ensure the daily operation of the system and reduce the unnecessary loss caused by system failure as far as possible (Fig. 1).

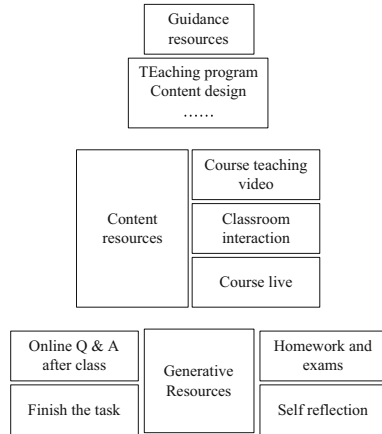


**Fig. 1.** Online teaching course route design

How to evaluate online courses has been discussed from different perspectives by many experts and scholars at home and abroad. For example, the MOOC education quality evaluation benchmark of the European Union of distance education universities is called “open education quality label”, which is mainly used to evaluate and compare the MOOC courses of schools. This set of evaluation mechanism includes two first level indicators, namely “institutional level” and “curriculum level”, which mainly correspond to 11 indicators, such as curriculum content, curriculum activities, curriculum structure, curriculum interaction and curriculum evaluation. When evaluating the quality of online teaching, there is a strong correlation between the indicators, which not only increases the workload, but also affects the effectiveness of the evaluation structure. Through the analysis and comparison of the evaluation methods of teachers’ teaching quality, we can easily think about the method of dimensionality reduction of indicators factor analysis. Dimensionality reduction reduces the original indicators to several main parts which no longer have linear relationship. The main factors obtained by factor analysis are usually difficult to have a specific meaning in real life, so we must rotate the principal components, and the rotated common factors have a clear meaning.

### 3.2 Evaluation Framework and Strategy

The design and development of online teaching quality reliability evaluation framework is an important part of the success of online teaching mode. It can also be said that the important factors of teaching success or failure are determined by curriculum resources. According to the different functions and uses of information-based teaching resources, our school classifies online course resources into three categories: guidance resources, content resources and generative resources (Fig. 2).



**Fig. 2.** Framework of online teaching course

#### Guiding Resources

It is not difficult to see from the figure above that the guiding resources are at the top of the pyramid. For the whole educational curriculum framework, they play a leading role and are programmatic resources. Combined with the characteristics of online teaching, the school puts forward the syllabus and teaching content which are different from the traditional teaching. Teachers write teaching materials according to the syllabus and content, so that students can quickly understand the objectives of online courses and get familiar with the contents of online courses. Due to the existence of big data technology, this stage is mainly teachers' teaching preparation, which includes the design of teaching scheme and the preparation of teaching resources. Before each class, teachers prepare courseware and videos suitable for online live teaching, and upload the general scope and outline materials of the class to the online education platform as pre class guidance materials. Push information from the platform to the student terminal to remind students. At this time, students can log on to the platform to watch these guiding resources, understand and be familiar with the content of the course, have a certain understanding of the knowledge about the upcoming course, and preview independently, so as to lay a certain foundation for the development of classroom teaching.



### **Content Resources**

It is located in the middle of the pyramid and plays a pivotal role as a link between the preceding and the following. At the same time, it is also the most important part of the whole teaching framework and the highlight of the teaching content. Curriculum resources are transferred from it to students. Once the design of content resources is defective, it will directly have a negative impact on the whole teaching framework. In the design of content-based resources, teachers generally use the idea of modularization to develop, summarize and divide the course content according to certain standards and connections, and modularize the teaching content, so that students can better understand and distinguish the learning content and learning objectives of each module [7–10]. This stage is the central link of the whole online teaching stage, the quality of classroom teaching directly affects the implementation of the whole teaching mode, and the content resources are the key to determine the success or failure of the classroom teaching stage. Unlike ordinary online teaching, which uses recorded video courses as course resources, the teaching in our school adopts the way of live video. The so-called teaching behavior should be a real-time process completed by teachers and students. There is a time gap between ordinary video teachers' recording and students' watching, which will cause further estrangement between teachers and students, It is not conducive to the development of teaching. At this stage, teachers' direct teaching is still the most efficient and direct teaching method. In the process of activities, people prefer to face a living person rather than a cold machine [1, 11, 12].

One of the advantages of online teaching is that it can carry out more communication between teachers and students. In order to ensure the quality of classroom teaching, students can be arranged to speak on behalf of each other after group discussion for many times in a semester. The students who speak can be scored [13–15]. When the final settlement of classroom performance, the system calculates the students who have not spoken or who have not spoken enough, the score of the project was discounted or even failed. At the same time, the students' attendance rate can be recorded through the student camera system. After the students log in to the system through the account, the system can monitor the students' performance in class. If there is no reason to leave, those who leave more times can be judged as late, leave early or even absent from class, which is more effective and accurate than taking time to call the roll. Du Juedai's attendance also saves time.

After the class assignment is finished, the teacher summarizes the class content and the students' speech, which is the end of the class teaching stage.

### **Generative Resources**

At the bottom of the pyramid, it is generated naturally from the use of the first two kinds of resources. In ordinary online courses, this part was once regarded as the closest and richest content in teaching practice, because there is no online live broadcast in ordinary online courses, and only when it comes to generative resources can there be interaction between teachers and students. But our school has expanded the content resources and advanced the interactive behavior of classroom teaching activities.

As the name suggests, this stage is the supplementary review after the completion of classroom teaching. Generally speaking, the teachers use the live broadcast software to set up the after-school question answering class at a fixed time. Although the online live broadcast class can let the students have no scruples about asking questions and discussing, after all, the classroom time is limited, so it is impossible to do everything. Therefore, increasing the communication time between teachers and students can help students better understand the course content, and review the subject content at the same time. Students can also log in to the system at any time during non question answering time to watch the video of online teaching course to consolidate their knowledge. The video content is temporarily reserved for one academic year. If they are rated as excellent demonstration course, the video content of the course will be further edited and launched as an online excellent course. If teachers have no special requirements, they usually submit the electronic version of homework through the system or email. After correcting, the teacher logs the scores into the system. The quality and quantity of students' homework in a semester can be clearly recorded, which is convenient for the final score scoring. In the final examination, our school combines 70% of the usual scores with 30% of the test paper scores. The usual scores can be obtained by referring to the class attendance rate, the number of classroom speeches and the scores of homework submission in the platform system, which is convenient for teachers to query and for the school to keep the bottom data.

### **3.3 Objective of Reliability Assessment**

In practice, with the support of big data technology, although the online education teaching performance evaluation method in China has begun to pay attention to speaking with data, it still relies on expert judgment and human input as a whole. For example, the general process of evaluating the teaching course of online education learning center by educational management institutions is to check teaching materials on site first, and then sample and interview teachers and students. For example, as a large-scale network higher education institution in China, the state Open University, namely, the state Open University, should conduct online teaching inspection on the nationwide school running institutions every year. Generally, it is to conduct spot check on the curriculum platform in various regions, organize dozens of experts to study the implementation of the sampled courses, and finally evaluate and score according to the indicators. The detection process is usually up to 4–5 months, At the same time, it consumes a large amount of human and material resources, so it can only be carried out once a year, and the sustainability and effectiveness of monitoring cannot be guaranteed. Moreover, the number of experts in the organization is limited, and the proportion of sampling courses is only a small part, and the comprehensive evaluation is also not guaranteed. As mentioned above, the data retrieval and analysis of online education platform teaching behavior, based on the data, establish an index system, and evaluate the results, which can be obtained with certain objectivity (Table 1).

**Table 1.** Configuration of online teaching evaluation platform

Processor	Intel Xeon processor e5-2660 V2
Operating system	Microsoft Windows Server 2012 R2 x64
Remote management	iDRAC8 Express
Power Options	Up to two 350 W hot swap redundant power supplies
Network adapter	Four port 1 GB Base-T adapter
Memory	32GB DDR3 DIMM
Storage	“3.5” enterprise sata7.2k hard disk x 4 (RAID10)
Computer room server group	CAT-6 cabling Gigabit Network
Customer terminal	Cat-5e wiring 100 m to desktop

#### 4 Analysis of System Practicability

To verify the design and application value of online teaching quality reliability evaluation model based on big data technology, the following comparative experiments are designed. 200 students in a university were selected as the experimental objects, of which 100 students were participants in the experimental group and 100 other students as the control group participants. For the former, online teaching quality reliability assessment model based on big data technology was adopted for the former curriculum quality assessment, and the traditional structural equation model was adopted for the latter (Fig. 3).

**Fig. 3.** Online teaching scene

The DSR (data set ready) data readiness index and the DUR (Data utilization durability) data utilization durability index can both reflect the reliability of the network host's evaluation of the quality of online teaching courses. In general, the physical values of the DSR index and the DUR index are more reliable. Larger, the higher the reliability of the web host's evaluation of the quality of online teaching courses, and vice versa, the lower. The following table records the specific numerical changes of the DSR index and DUR index of the experimental group and the control group.

**Table 2.** DSR index values

Experimental time/(min)	DSR index value/(%)	
	Experience group	Control group
5	67.3	40.5
10	68.1	40.5
15	69.6	40.4
20	70.2	40.3
25	70.5	40.2
30	70.7	40.1
35	71.0	40.0
40	71.4	39.9
45	71.8	39.8
50	72.2	39.8

Analyzing Table 2 shows that with the extension of the experiment time, the DSR index of the experimental group always maintained a rising value trend; while the DSR index of the control group maintained a first stable, then decreased, and finally stable value trend. From the perspective of the limit value, the maximum value of 72.2% in the experimental group is an increase of 31.7% compared with the maximum value of 40.5% in the control group.

**Table 3.** DUR index values

Experimental time/(min)	DUR index value/(%)	
	Experience group	Control group
5	63.6	32.1
10	63.6	32.3
15	63.8	32.5
20	63.8	32.7
25	64.1	32.9
30	64.1	33.2
35	64.3	33.4
40	64.3	33.6
45	64.7	33.6
50	64.7	33.6

Analysis of Table 3 shows that with the extension of the experiment time, the DUR indicator of the experimental group has always maintained a numerical trend of rising in a stepped manner; the DUR indicator of the control group has gradually tended to a relatively stable numerical performance after a period of numerical upward trend. From the perspective of the limit value, the maximum value of 64.7% in the experimental group is 31.1% higher than the maximum value of 33.6% in the control group.

In summary, with the application of the online teaching quality reliability evaluation model based on big data technology, the value of DSR index and DUR index have shown a significant upward trend, which can fundamentally realize the effect of the network host on the quality of online teaching courses. Reliability assessment.

## 5 Concluding Remarks

With the support of big data technology, online teaching quality evaluation involves different disciplines, majors, and teachers. The sample selected in this empirical study is the index data of a university, which requires higher reliability and effectiveness. For the verification analysis, it may be necessary to find more suitable samples and interviewees. In addition, this article only uses the measured samples to train the model, and obtains a model structure with reliability analysis. If you want to prove the wide applicability of the model, you can rely on the structure relationship to calculate the weight and use the regression method to solve the final results of each discipline. The ranking order is compared with the school evaluation results. In the future, the model can be checked and accepted with the data of other majors in other colleges and universities.

## References

1. Li, Z., Wu, W., Tai, X., et al.: Optimization model-based reliability assessment for distribution networks considering detailed placement of circuit breakers and switches. *IEEE Trans. Power Syst.* **35**(5), 3991–4004 (2020)
2. Miao, Y.H., Zuo, P.P., Yin, J., et al.: An improved CPTu-based method to estimate jacked pile bearing capacity and its reliability assessment. *KSCE J. Civ. Eng.* **23**(9), 3864–3874 (2019)
3. Chou, P.H., Chiang, K.N., Liang, S.Y.: Reliability assessment of wafer level package using artificial neural network regression model. *J. Mech.* **35**(6), 1–9 (2019)
4. Liu, X., Liu, Z., Chen, J.T., et al.: Ergonomic reliability assessment for passenger car interface design based on EWM-MADM and human cognitive reliability experiments. *Math. Probl. Eng.* **20**(5), 1–10 (2020)
5. Zheng, P., Shuai, L., Arun, S., Khan, M.: Visual attention feature (VAF): a novel strategy for visual tracking based on cloud platform in intelligent surveillance systems. *J. Parallel. Distrib. Comput.* **12**(5), 182–194 (2018)
6. Meng, Z., Zhang, Z., Zhou, H.: A novel experimental data-driven exponential convex model for reliability assessment with uncertain-but-bounded parameters. *Appl. Math. Model.* **77** (11), 773–787 (2020)
7. Zhang, L., Jin, G., You, Y.: Reliability assessment for very few failure data and Weibull distribution. *Math. Probl. Eng.* **19**(4), 1–9 (2019)
8. Liu, S., Bai, W., Zeng, N., Wang, S.: A fast fractal based compression for MRI images. *IEEE Access* **7**(2), 62412–62420 (2019)
9. Kozubal, J.V., Pua, W., Stach, M.: Pile in the unsaturated cracked substrate with reliability assessment based on neural networks. *KSCE J. Civ. Eng.* **23**(1), 1–11 (2019)
10. Rahim, A., Abdullah, S., Singh, S., et al.: Reliability assessment on automobile suspension system using wavelet analysis. *Int. J. Struct. Integrity* **10**(5), 602–611 (2019)

11. Mengye, L., Liu, S., Kumarsangaiah, A., et al.: Nucleosome positioning with fractal entropy increment of diversity in telemedicine. *IEEE Access* **6**, 33451–33459 (2018)
12. Chen, F., Li, F., Feng, W., et al.: Reliability assessment method of composite power system with wind farms and its application in capacity credit evaluation of wind farms. *Electric Power Syst. Res.* **166**(21), 73–82 (2019)
13. Liu, P., Wang, X., Teng, F.: Online teaching quality evaluation based on multi-granularity probabilistic linguistic term sets. *J. Intell. Fuzzy Syst.* **15**(2), 1–20 (2021)
14. Yu, H.: Online teaching quality evaluation based on emotion recognition and improved AprioriTid algorithm. *J. Intell. Fuzzy Syst.* **40**(5), 1–11 (2020)
15. Yang, Y.D.: An online evaluation method on the effect of flipping task-based English reading teaching. *J. Bengbu Univ.* **8**(04), 77–81 (2019)



# Reliability Detection Method of Online Education Resource Sharing Based on Blockchain

Feng Wang<sup>(✉)</sup>

Jiujiang University, Jiujiang 332005, China  
wangfengl7172021@163.com

**Abstract.** In order to reduce the error of online education resource sharing reliability detection, a block chain based online education resource sharing reliability detection method is designed. This paper first analyzes the status of online education resource sharing, then defines the reliability of information resource sharing system, and finally establishes the reliability detection model of online education resource sharing to realize the reliability detection of education resource sharing. The experimental results show that the proposed method has smaller detection error and higher sharing security than traditional methods.

**Keywords:** Blockchain · Online education · Resource sharing · Reliability

## 1 Introduction

With the development and popularization of computer network technology and multimedia technology, network education, as a form of education, has developed rapidly in higher vocational colleges. It breaks through the limitation of time and space in the traditional education process and realizes the sharing of teaching resources in a wider range. At present, many countries attach great importance to the construction of online teaching resources, such as the “educational technology action plan” and “American Education Action Plan” of the United States, the CTI plan of the United Kingdom, the school networking plan of France, the education network plan of Australia, the “education revolution” of South Korea, the cross century plan of Singapore and the school networking experiment of Japan. However, there are many problems in the security of campus network, so the security of network teaching resources can not be effectively guaranteed, which also restricts the construction and sharing of network teaching resources in Higher Vocational Colleges to a certain extent.

Any science and technology is a “double-edged sword”. Information resource sharing system has its own vulnerability. System security and reliability is the most important and basic requirement, and also the premise for the existence and development of information resource sharing system. How to improve the reliability of the system and prevent the unreliable factors is one of the keys of the system design. Therefore, the reliability detection method of online education resource sharing is studied. Traditional methods generally use data mining to share online education

resources, but the reliability of sharing is poor because of the huge amount of online education resources data and the huge types of resources and data involved.

Blockchain is a kind of database with hash verification function. Block is the data block. It combines the data blocks into a chain structure according to the time sequence, and uses the cryptography algorithm to maintain the reliability of the database collectively in the way of Distributed Accounting. All data blocks are connected in chronological order to form a blockchain. Based on the characteristics of blockchain technology, this paper designs an online education resource sharing reliability detection method based on blockchain. On the basis of clarifying the status of online education resource sharing, the reliability of education resource information sharing system is defined. Based on this, the reliability testing model of online education resource sharing is established to realize the reliability testing of education resource sharing.

## **2 Online Education Resource Sharing**

In today's global education information environment, the construction of information-based education environment is an important part of education modernization. In the current higher vocational colleges, network teaching resources have the advantages of timeliness of information and resource transmission, students' autonomy in learning, interaction between teaching and learning, and sharing of teaching resources. All schools attach great importance to it, and build their own network teaching resource system to reform the traditional teaching and management mode, and build a network learning platform for students' autonomous learning and collaborative learning [1]. At present, in the general campus network, network teaching resource system is mainly divided into four layers: basic layer, resource management layer, application layer and user layer. Among them, the basic layer refers to the basic platform of campus network operation, and CERNET and Chinanet are the main connecting networks of higher vocational colleges; The resource management layer is generally composed of database server and resource server group, such as IP server, e-mail server and streaming media server; The application layer is mainly composed of web server and related applications; The user layer is mainly composed of teachers and students in the campus network. The system structure of the four layer network teaching resource management system has laid a solid foundation for the sharing of teaching resources in campus network, but there are many security problems in the actual use of network teaching resources, which will restrict the construction and development of network teaching resource system in higher vocational colleges. Therefore, it is necessary to analyze these security problems and find a reasonable method of sharing network teaching resources.

## **3 Reliability Definition of Information Resource Sharing System**

In the actual operation, information resource sharing system may face various errors or exceptions, which can be attributed to some external factors (such as artificial error, network or hardware abnormality). An information resource sharing system should



ensure that there is a pre-defined framework or mechanism to handle errors or recover from errors. Therefore, reliability becomes an important index of the performance of distributed information resource sharing system.

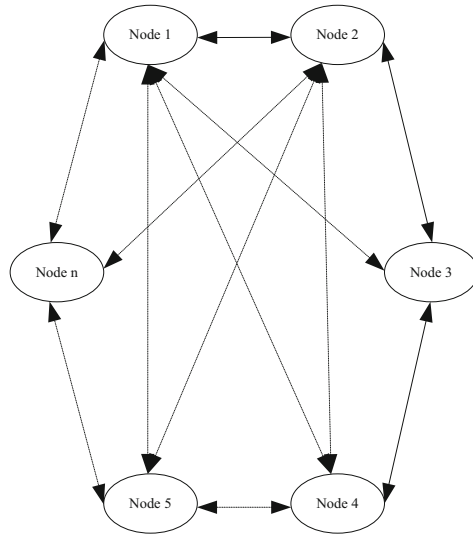
Reliability refers to the probability that the system operates normally and correctly performs the required functions under the specified operating environment and within the specified time period. The strong reliability of the system means that when the external environment or internal state changes or abnormally, the system can still complete its preset function, without stopping the system running or crashing. Information resource sharing system is a highly complex information system, which requires better flexibility. Dynamic, distributed and intelligent systems are the best choice to support information resource sharing. However, such a system has high uncertainty, so it is necessary to give the reliability evaluation of information resource sharing system. The reliability of the product refers to the ability of the product to complete the specified functions under the specified conditions and within the specified time (or operation times). The reliability of products is similar to that of information resource sharing system. The following definitions are given for reliability of information resource sharing system:

Definition 1: the reliability of information resource sharing system refers to the ability of information resource sharing system to complete specified tasks under specified conditions and within specified time.

Definition 2: the reliability of information resource sharing system refers to the probability that the information resource sharing system can complete the specified functions under the specified conditions and within the specified time.

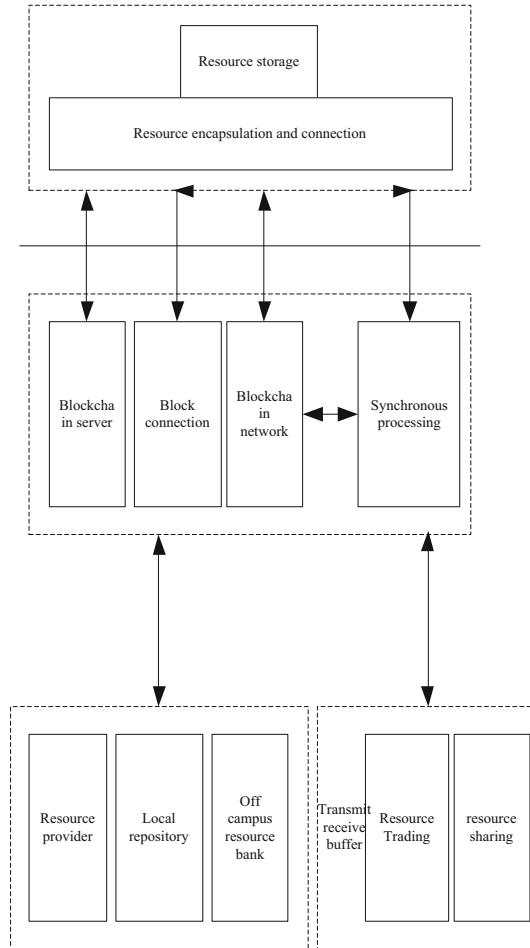
## **4 Establishment of Education Resource Database Based on Blockchain**

According to the different transmission speed and form of educational resources, educational resources sharing can be divided into two modes: centralized mode and non centralized mode; Resource management without central mode takes decentralized supervision as its operation mode, and mainly uses point-to-point technology to realize sharing mode, which has the characteristics of convenient management and high data security [2–6]. On the basis of fully considering the respective advantages of the two modes, this paper constructs a network structure of digital education resource database with centralized supervision and no central structure, as shown in Fig. 1.



**Fig. 1.** Teaching resource database structure based on blockchain Technology

Among them, the nodes of the blockchain are sequential, and the node serial number is one of the basic attribute information of the block. In the framework of digital education resource sharing network structure, the main body of each node can be school teachers, training institutions or learners, and the nodes can be connected with each other. At the same time, in the framework model of digital education resources, blockchain technology is used as the underlying framework, and consensus mechanisms such as encryption algorithm, rule verification, digital signature and education smart contract system with high technical maturity are used to realize specific operations such as automatic loading, downloading and updating of education resources. Therefore, based on the principle of blockchain, this paper constructs a dynamic extensible, manageable, controllable and open service blockchain digital education resource database model, as shown in Fig. 2.



**Fig. 2.** Database structure

The blockchain digital education database model mainly includes four layers. ① Resource storage layer. First of all, we need to form books, videos, files and other educational resources into digital resources and encapsulate them in the form of blocks, and send them to local resources (servers). ② Resource connection layer. The packaged block resources form a block interconnection group through consensus mechanism and encryption algorithm, and are synchronized and updated in the network resource library. ③ Resource transaction layer. The resource provider encapsulates the digital resource in the block waiting for verification [7–15] (the basic data required in the block is shown in Table 1).

**Table 1.** Basic data in block

Serial number	Block data type	Explain
1	Node number	Node serial number is also the basic attribute information of block
2	Version number	Identification number of block version
3	Block size	Bytes representing data
4	Time stamp	A sequence of characters that uniquely identifies the time of a moment
5	Block address value	It is stored in block head, and is mainly obtained by hash algorithm SHA-256
6	Merkel root	A hash binary tree, invented by Ralph Merkle in 1979
7	Other information	Block related information, including information of resource provider, etc.

Then, the system connects each block to the main chain of the previous block in the blockchain system, and backs up the data information in each blockchain to all nodes, that is, each node records all the data in the whole blockchain. The block contains basic data information such as version number, time stamp and block address value, which makes clear the ownership of digital education resources. In case of infringement, each node can act as a “witness”. ④ Resource sharing layer. The local resource database and the off campus resource database adopt the tree topology structure to save the data information of each node, so as to further improve the query speed of data information. As can be seen from Fig. 1 and Fig. 2, the digital education resource database established in this paper is a B2B network spontaneously formed by each node according to the demand for education resources. Learners, school teachers and training institutions can become a specific node in the blockchain network system after a series of security verification such as encryption algorithm, rule verification and digital signature. Network node is the core foundation of system operation, which is mainly responsible for generating blocks and forming blockchain network through hash algorithm connection. The specific connection process is shown in Fig. 3.

It can be found that the block head and block body constitute the basic structure of the block. Among them, the block head mainly includes the root value of Merkel tree, time stamp, difficulty value and other specific basic data information, and the block body usually contains multiple transaction records.

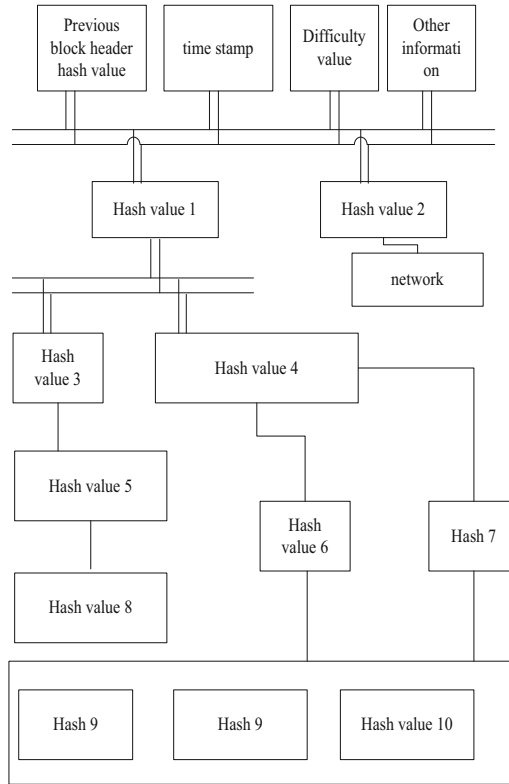


Fig. 3. Block structure of connection system

## 5 Reliability Evaluation Model of Educational Resource Sharing

### 5.1 Information Denoising

If there is noise in the reliability signal of network resource sharing system, it will have a certain impact on the evaluation results. It is necessary to denoise the reliability signal of the system. The following is the specific research process: the continuous form of wavelet function is given by using formula (1):

$$\Phi_{a,b}(t) = |a|^{-M/2} \Phi\left(\frac{t-b}{a}\right) \tag{1}$$

In the formula,  $a$  is the correlation coefficient of wavelet coefficients,  $b$  is discretized,  $\Phi$  is scale parameter,  $M$  is denoising parameter.

## 5.2 Determination of Risk Assessment Method

Risk analysis is a method that teaching resource sharing system can systematically assess the risk of information security in the process of operation, and determine the level of risk based on it. It mainly includes qualitative analysis and quantitative analysis (including semi quantitative analysis) as shown in Table 2. Considering the complexity and cost [16–20], both qualitative analysis and quantitative analysis are very high, and quantitative analysis is more than qualitative analysis.

**Table 2.** Risk assessment method

Serial number	Method	Primary coverage
1	Qualitative analysis	Use text or descriptive data ranges to describe the magnitude of potential risks and the likelihood of occurrence of these risks
2	Semi quantitative analysis	In semi-quantitative analysis, the numerical range of qualitative analysis is known value, and the number referred to in each description may not be able to accurately represent the actual degree of risk influence semi-quantitative analysis or possibility. The purpose of semi-quantitative analysis is to obtain a more detailed degree of risk than qualitative analysis, but does not need to propose any actual value of risk obtained in quantitative analysis
3	Quantitative analysis	Quantitative analysis uses numerical value in the analysis of influence or possibility, while non quantitative analysis is the narrative numerical range applicable to qualitative or semi quantitative analysis, and uses data obtained from various sources and channels. The quality of quantitative analysis depends on the accuracy and integrity of the data used

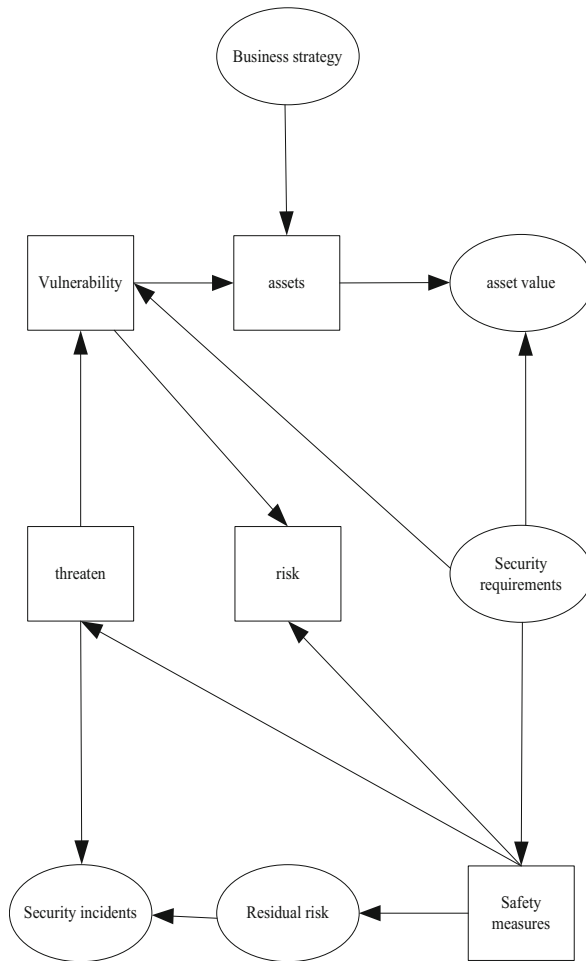
At present, the risk assessment of teaching resource sharing system is mostly entrusted to security product manufacturers. It is difficult to determine the security requirements such as what to protect, where to protect the boundary of the object, and how much to protect. The establishment of evaluation methods must take into account the conditions and regulations of the teaching resource sharing system, such as the established scope of information security management system, information security implementation requirements, laws and regulations to be followed in the operation process, so as to achieve the effect and improve the efficiency. Based on the risk assessment of their own safety, a suitable quantitative analysis method in the process of being selected is mainly based on the evaluation of the effect, workload, cost and benefit, technical complexity and the difficulty of data collection.

### Identification and Evaluation of Reliability

Within the scope of information security management system, teaching resource sharing system should be able to identify the assets, the principals, the elements of threatening assets, which weaknesses may be threatened to be used, and the potential impact on assets when availability, confidentiality or integrity are lost. Assess the security failures that may lead to commercial impact based on the potential impact of identified assets; Based on the main threats, weaknesses and impacts of assets, and the control measures being implemented, the realistic possibility of failure caused by such causes is evaluated; According to the established risk level criteria, the risk level is determined.

### The Relationship Between the Elements of Reliability Evaluation

The work of reliability assessment must focus on the basic elements of assets, threats to information security, vulnerability in the process of information security implementation, security measures and risks. In the evaluation process of these elements, we must give full consideration to the related attributes, including the implementation of business development strategy, asset value, the events that affect the security in the development process, and the residual risks after. The relationship between the relevant elements in the risk assessment is shown in the figure, and the terms are explained in the table. In the figure, the basic element content is represented in the box, and the attribute content is represented in the ellipse. The relationship among the elements of reliability evaluation is shown in Fig. 4 below.



**Fig. 4.** Relationship of various elements of risk assessment

Table 3 below is a glossary of terms:

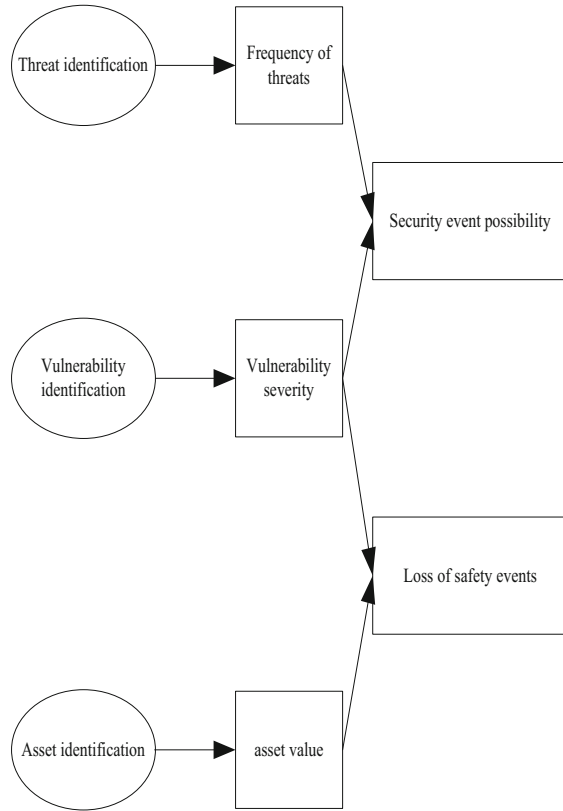
**Table 3.** Glossary

Serial number	Name	Primary coverage
1	Business strategy	The tasks of an organization realized by means of information technology
2	Assets	Anything of value to the unit
3	Asset value	Importance and sensitivity of assets
4	Threaten	Potential causes of accidents that may cause damage to assets or units
5	Risk	Due to the vulnerability of the system, the possibility and impact of security incidents caused by man-made or natural threats
6	Residual risk	After the safety protection measures are taken and the protection ability is improved, the possible risks still exist
7	Security incidents	If the threat subject can produce a threat and make use of the vulnerability of assets and their security measures, then the actual harm situation is called a security event

### Reliability Evaluation Model Construction

Each element in the reliability evaluation model has its own unique attributes. In the process of calculating the reliability value, the first thing is to know the value of assets; The vulnerability of information assets is identified and its severity is assigned; Analyze the threat and its frequency; Fourth, the probability of security incidents is calculated according to the frequency and vulnerability of threats; Finally, the reliability is calculated according to the importance of information assets and the probability of security events on the information assets. The qualitative analysis of vulnerability is basically that the suppliers of products provide non teaching resource sharing system, which will hinder the determination of security requirements. The frame is shown in Fig. 5 below.





**Fig. 5.** Schematic diagram of risk assessment model

Reliability is one of the key factors in the design, research and operation of information resource sharing system. Due to the limitation of resources, such as cost, reliability and time, the optimization of system reliability has attracted extensive attention. For this reason, taking three optimization objectives of reliability, cost and time as examples, the multi-objective mathematical programming of workflow system reliability optimization is established, as shown in the following formula:

$$\left. \begin{aligned}
 & \max R(R_1, R_2, \dots, R_N) \\
 & \min C(R_1, R_2, \dots, R_N, T_1, T_2, \dots, T_N) \\
 & \min T(T_1, T_2, \dots, T_N) \\
 & s.t. R(R_1, R_2, \dots, R_N) > R_{\min} \\
 & C(R_1, R_2, \dots, R_N) > C_{\max} \\
 & T(T_1, T_2, \dots, T_N) > T_{\max} \\
 & R_{i,\min} < R_i \cdot R_{i,\max}
 \end{aligned} \right\} \quad (2)$$

In formula (2), the reliability of system  $R$  and the cost of system  $C$  are also functions of the reliability and time of each unit;  $R_{\min}$  is the lower bound of system reliability;  $C_{\max}$  is the upper bound of system reliability;  $T_{\max}$  is the upper bound of system time;  $R_{i,\max}$ . The upper and lower limits of the reliability of  $R_{i,\min}$  element are constrained.

At the same time, in the actual network environment, the failure probability of the platform link and node is random and independent, and has known failure probability. The evaluation method of service platform reliability is mainly used to study whether the platform signal is stable and reliable. In the practical application network, for the irreparable information consulting service platform, the platform signal experiences the process of reliability decline until the communication of each signal fails. For repairable system, it will automatically enter the repair state after failure.

With the improvement of the reliability of hardware components in service platform. The average time of failure free is also increasing, CPU, memory and so on have a certain life cycle. When these hardware components are integrated to form communication links and nodes of the platform, the MTBF of the reliability signal of the consulting service platform is expressed as:

$$R_s = \left( \sum_N^{k=1} \frac{1}{R_k} \right)^{-1} \tag{3}$$

In formula (3),  $R_k$  represents the MTBF of component  $K$  in the service platform, and  $N$  represents the number of components in the service platform. Intuitively, the most unreliable component in the service platform determines the quality of the reliability signal of the whole service platform. Although the reliability of the whole platform is constantly improving, the integration of the platform is also increasing. A large number of platform components will lead to the decline of platform signal reliability.

Before the service platform is officially used, it is necessary to select qualified service platform components after the probation period and aging period. According to the energy description chart of service platform failure rate, the Weibull function is used to represent the failure rate curve  $\alpha(t)$ :

$$\alpha(t) = \begin{cases} 0, & 0 \leq t \leq t_0 \\ \frac{a}{b}(t+c-t_0)^\alpha, & t_0 \leq t \leq t_2 \\ \alpha, & t_1 \leq t \leq t_2 \\ \frac{a}{b}(t-t_2)^\alpha + \alpha, & t_2 \leq t \end{cases} \tag{4}$$

In formula (4),  $a$ ,  $b$  and  $c$  represent different shape coefficients. By measuring the failure rate of platform reliability signal at any time by actual failure measurement rate, the following failure rate is obtained at time  $t_1$ :

$$\alpha = \frac{a}{b} c^\alpha \tag{5}$$

In the actual network, the node and link efficiency of information consulting service platform mainly refers to the ratio of the current ability of node and link to complete the platform task to its maximum efficiency. In a non repairable system, the performance is expressed as follows:

$$E(t) = \exp\left(-\int_0^t \alpha(t)dt\right) \tag{6}$$

The network model is used to analyze the current network efficiency, and the premise of calculating network efficiency is to ensure the effectiveness of the network. Network efficiency is obtained by using all nodes and link efficiency in the network through topology performance, then network performance  $E_N(t)$  is expressed as:

$$E_N(t) = E_e(t)E_v(t) \tag{7}$$

In formula (7),  $E_v(t)$  represents the total node efficiency of the network resource sharing information consulting service platform, and the network efficiency is to ensure that each node in the network is effective. If the performance of a node in the network is 0, then  $E_v(t)$  can be expressed as:

$$E_v(t) = \prod_n^{i=l} E_{vi}(t) \tag{8}$$

In the above formula (8),  $E_{vi}(t)$  represents the efficiency of node  $c$ , and  $n$  represents the number of nodes of the network resource sharing information consulting service platform.

**The Solution of the Model**

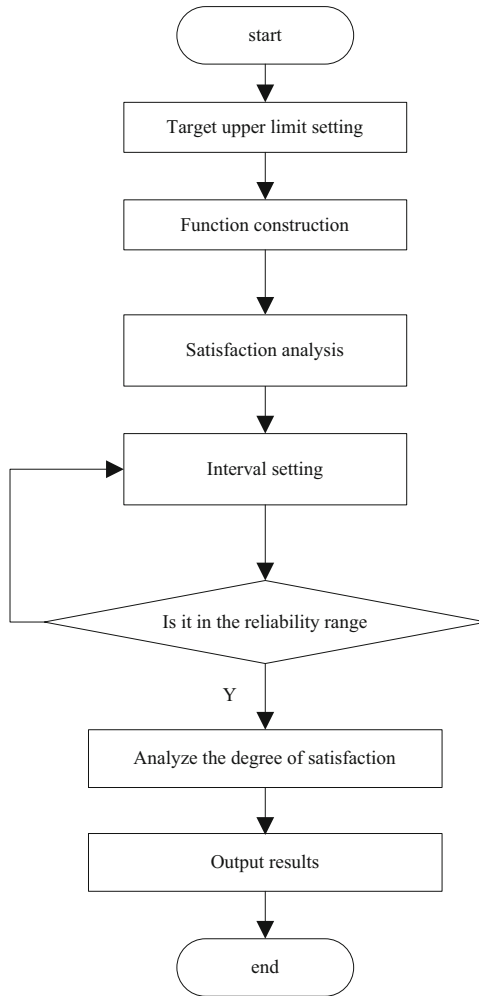
The target lower limit  $R_{\min}$  for the reliability of information resource sharing system is given. The target upper limit  $R_{\max}$  of system reliability can be obtained by the following optimization model of maximum reliability which satisfies the cost constraints:

$$\left. \begin{aligned} &\max R(R_1, R_2, \dots, R_N) \\ &s.t. C(R_1, R_2, \dots, R_N) > C_{\max} \\ &T(T_1, T_2, \dots, T_N) > T_{\max} \\ &R_{i,\min} < R_i \cdot R_{i,\max} \\ &T_i > 0 (i = 1, 2, \dots, N) \end{aligned} \right\} \tag{9}$$

The target upper limit of information resource sharing system cost has been given. The target lower limit of system cost can be obtained by the following formula:

$$\left. \begin{aligned}
 &\min C(R_1, R_2, \dots, R_N, T_1, T_2, \dots, T_N) \\
 &s.t. R(R_1, R_2, \dots, R_N) > R_{\min} \\
 &T(T_1, T_2, \dots, T_N) < T_{\max} \\
 &R_{i,\min} < R_i < R_{i,\max}
 \end{aligned} \right\} \quad (10)$$

The objective satisfaction function takes the objective function as the independent variable, and the target satisfaction represents the satisfaction degree of the decision-maker with the value of the objective function. The value of 1 indicates the most satisfied and the value of 0 indicates the most dissatisfaction. The specific process is shown in Fig. 6 below:



**Fig. 6.** Satisfaction calculation process

In general, it can be considered that in the system reliability tolerance interval  $[R_{\min}, R_{\max}]$ , the information resource sharing system reliability is the most satisfactory with the upper limit value  $R_{\max}$ , that is, the satisfaction degree of  $R_{\max}$  is 1; And the reliability of information resource sharing system is the lowest limit,  $R_{\min}$  is the most dissatisfied, that is, the satisfaction of  $R_{\min}$  is 0.

The satisfaction function  $h_1(R)$  of the reliability of information resource sharing system is a monotonic increasing function with the value in the interval  $[0, 1]$ . generally, it should be in the form of the following Fig. 7:

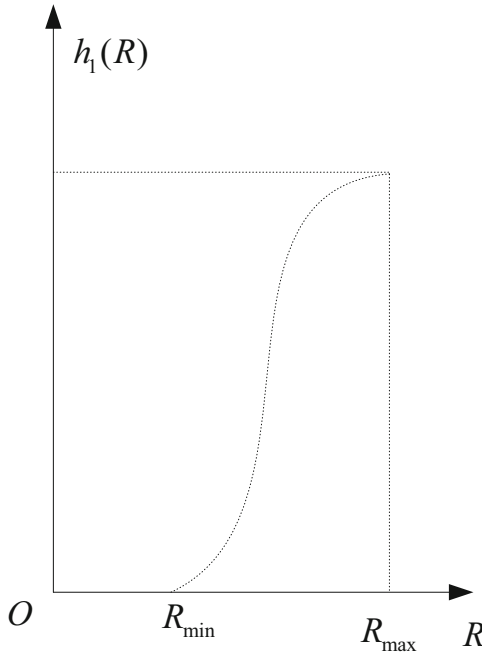


Fig. 7. Reliability satisfaction function

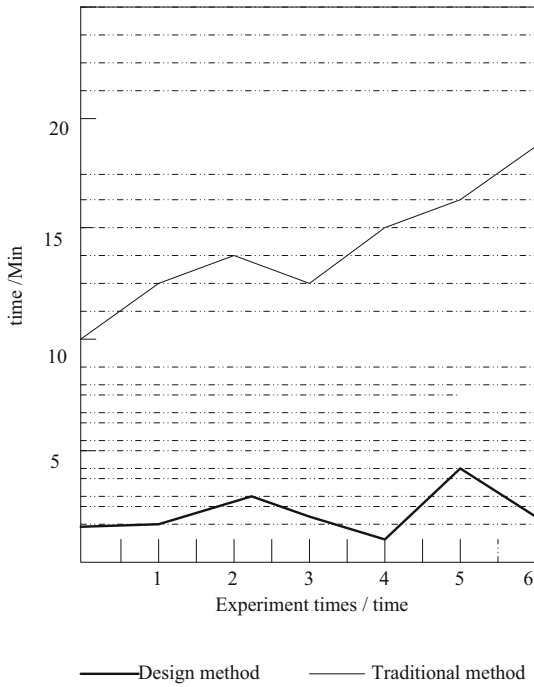
Through the above process, the reliability of online education resources sharing is detected.

## 6 Experimental Comparison

In order to verify the comprehensive effectiveness of the reliability detection method of online education resource sharing based on blockchain, simulation experiments are needed. Experimental environment: Visual Studio 2008 platform under win7 system, Intel Xeon processor, 4 GB memory, NVIDIA Quadro fx1800 display card. And the traditional detection method and the research method are compared to compare the detection effect of the two methods.

### 6.1 Comparison of Detection Efficiency

The comparison results of detection efficiency of the two methods are shown in Fig. 8.

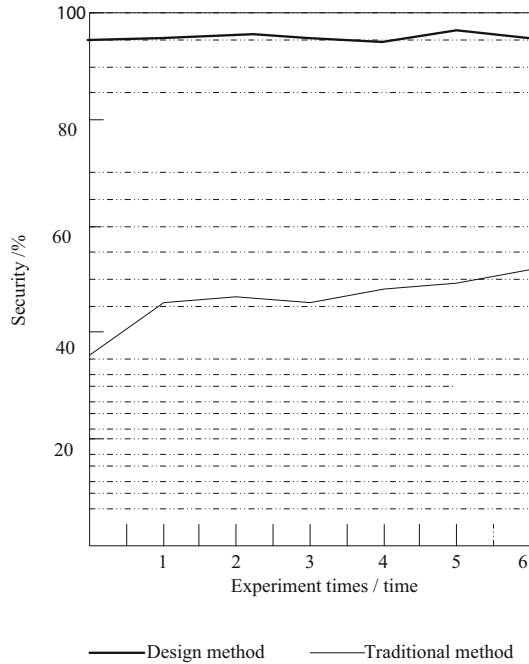


**Fig. 8.** Comparison of detection efficiency

Through the analysis of the figure above, it can be seen that the detection time of the reliability detection method in this study is less than that of the traditional method, which proves that the method in this study has higher detection efficiency.

### 6.2 Security Comparison of Educational Resources Sharing

The comparison results of the security of educational resource sharing between the traditional method and the studied method are shown in Fig. 9 below.



**Fig. 9.** Security comparison of educational resources sharing

Through the analysis of the above content, it is found that the sharing security result of this research method is higher than that of traditional methods, which has a certain practical significance.

## 7 Concluding Remarks

To sum up, the reliability detection method in this study can overcome the vulnerability of information resource sharing system, quantitatively analyze the reliability of information resource sharing system, improve the reliability of system, and prevent unreliable factors. It can provide a scientific basis for improving the security and reliability of the system, and the related theories, sharing models and applicable software development of information resource sharing system still need to be further studied.

## References

1. Tan, G., Straub, N., Kaczmarek, S., Henke, M.: Blockchain-technologie im interdisziplinären umfeld. *ZWF Zeitschrift fuer Wirtschaftlichen Fabrikbetrieb* **114**(10), 605–609 (2019)
2. Rasool, S., Saleem, A., Iqbal, M., Dagiuklas, T., Qayyum, Z.U.: Docschain: blockchain-based IoT solution for verification of degree documents. *IEEE Trans. Comput. Soc. Syst.* **7** (3), 827–837 (2020)

3. Hameed, S., Shah, S.A., Saeed, Q.S., Siddiqui, S., Draheim, D.: A scalable key and trust management solution for iot sensors using SDN and blockchain technology. *IEEE Sens. J.* **21**(6), 8716–8733 (2021)
4. Kim, H.M., Laskowski, M., Zargham, M., Turesson, H., Kabanov, D.: Token economics in real life: cryptocurrency and incentives design for insolar's blockchain network. *Computer* **54**(1), 70–80 (2021)
5. Navy, S.L., Nixon, R.S., Luft, J.A., Jurkiewicz, M.A.: Accessed or latent resources? Exploring new secondary science teachers' networks of resources. *J. Res. Sci. Teach.* **57**(2), 184–208 (2020)
6. Han, J., Kim, D.: Security offloading network system for expanded security coverage in ipv6-based resource constrained data service networks. *Wirel. Netw.* **26**(6), 4615–4635 (2020)
7. Fan, P., Liu, Y., Zhu, J., Fan, X., Wen, L.: Identity management security authentication based on blockchain technologies. *Int. J. Netw. Secur.* **21**(6), 912–917 (2019)
8. An, Y., Xu, M., Shen, C.: Classification method of teaching resources based on improved knn algorithm. *Int. J. Emerg. Technol. Learn. (iJET)* **14**(4), 73–88 (2019)
9. Yi, B., Wang, X., Huang, M., Yang, L.: Cost and security-aware resource allocation in optical data center networks. *IEEE Commun. Lett.* **23**(11), 2031–2035 (2019)
10. Yang, W., Zhao, X., He, J.: Physical layer security and energy efficiency driven resource optimisation for cognitive relay networks. *IET Commun.* **14**(17), 2953–2961 (2020)
11. Kibiwott, K.P., Zhang, F., Kimeli, V.K., Anyembe, O.A., Opoku-Mensah, E.: Privacy preservation for ehealth big data in cloud accessed using resource-constrained devices: survey. *Int. J. Netw. Secur.* **21**(2), 312–325 (2019)
12. Bai, X., Li, J.: Intelligent platform for real-time page view statistics using educational big data digital resource sharing. *J. Intell. Fuzzy Syst.* **40**(1), 1–10 (2020)
13. Yuan, Q.: Network education recommendation and teaching resource sharing based on improved neural network. *J. Intell. Fuzzy Syst.* **39**(4), 5511–5520 (2020)
14. Wittmann, M.C., Millay, L.A., Alvarado, C., Lucy, L., Rogers, A.: Applying the resources framework of teaching and learning to issues in middle school physics instruction on energy. *Am. J. Phys.* **87**(7), 535–542 (2019)
15. Liu, S., Bai, W., Zeng, N., Wang, S.: A fast fractal based compression for MRI images. *IEEE Access* **7**, 62412–62420 (2019)
16. Pender, L.K., Kadkhoda, H., Lucero, K.S., Repetto, P., Weber, J.S.: Effects of online education on the identification and management of immune-related adverse events over time. *J. Clin. Oncol.* **37**(15), 18224 (2019)
17. Liu, S., Pan, Z., Cheng, X.: A novel fast fractal image compression method based on distance clustering in high dimensional sphere surface. *Fractals* **25**(4), 1740004 (2017)
18. Dai, Y., Xu, D., Maharjan, S., Chen, Z., He, Q., Zhang, Y.: Blockchain and deep reinforcement learning empowered intelligent 5g beyond. *IEEE Netw.* **33**(3), 10–17 (2019)
19. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–17 (2018). <https://doi.org/10.1007/s11036-018-1134-8>
20. Luo, B., Li, X., Weng, J., Guo, J., Ma, J.: Blockchain enabled trust-based location privacy protection scheme in vanet. *IEEE Trans. Veh. Technol.* **69**(2), 2034–2048 (2020)





# Research on Inventory Control Method of Cold Chain Logistics Enterprises Under the Background of New Energy Consumption

Dong-ming Yue<sup>1(✉)</sup>, Hui-ling Zhang<sup>1</sup>, Guang-yao Miao<sup>1</sup>,  
Le-le Wang<sup>1</sup>, and Feng-shuo Yan<sup>2</sup>

<sup>1</sup> Marketing Service Center (State Grid Ningxia Electric Power Co., Ltd. Metering Center), State Grid Ningxia Electric Power Co., Ltd., Ningxia 750000, China

yuedongming@nx.sgcc.com.cn

<sup>2</sup> The Second Research Institute of Civil Aviation Administration of China, Chengdu 610041, China

**Abstract.** The conventional inventory control method does not accurately calculate the inventory cost of cold chain logistics, which leads to its poor application effect. In order to optimize the inventory control capability, this paper designs a new inventory control method for cold chain logistics enterprises based on the background of new energy consumption. The cold chain logistics inventory cost is calculated by establishing the cold chain logistics inventory cost model, and the ordering cost, inventory holding cost, shortage loss cost, transportation cost and delay time cost of manufacturers, suppliers and distributors are obtained respectively. Then, the dynamic model of cold chain logistics inventory control is built, and the algorithm of cold chain logistics inventory control is designed. The experiment shows that the application effect of this method is better than that of three conventional methods, and it proves that it can optimize the inventory control process of cold chain logistics enterprises.

**Keywords:** New energy · Background of dissipation · Cold chain logistics enterprises · Inventory control method

## 1 Introduction

Cold chain logistics inventory problem is one of the key problems in the whole cold chain development process. For the enterprises with all links in the whole cold chain, from the origin suppliers, product manufacturers to sellers, improving the inventory control of the cold chain can reduce the inventory level of the enterprise, and the saved investment can be used in other links of the working capital, while meeting the changing needs of consumers, Minimize the adverse effects of inventory shortage and product deterioration and waste caused by inventory overstock.

China's cold chain products have many operation links, complex and diverse circulation channels, and the cold chain system is not perfect, resulting in serious loss of cold chain products in the circulation process [1]. The lagging development of cold storage can easily lead to deterioration, decay and waste of products. By studying the

cold chain inventory control model and seeking optimization, it is of great theoretical and practical significance to reduce logistics costs and avoid unnecessary waste.

In reference [2], by establishing a dynamic model of reinforcement learning in cold chain management, an inventory control method which can minimize the logistics cost is proposed to solve the problem of mutual inhibition among various factors. Reference [3] uses SWOT analysis method to solve the problem of macro and micro overall analysis dislocation, points out the uneven demand and supply in the logistics process, and puts forward an effective solution, which integrates comparative analysis method into inventory control method. Reference [4] analyzes the current situation of cold chain consumption on the basis of macro-control, further traces the specialized and refined cold chain management system, designs the relevant network architecture, and designs a multi-level coordinated inventory control method. However, in practical application, it is found that the practical application effect of the above methods is not ideal.

Aiming at the shortcomings of traditional methods, this paper designs a new inventory control method for cold chain logistics enterprises based on the background of new energy consumption, and combines new energy distribution with logistics to form a feasible inventory plan. The main ideas of the method in this paper are as follows:

- (1) Establish the cold chain logistics inventory cost model, and use this model to calculate the cold chain logistics inventory cost, so as to obtain the ordering cost, inventory holding cost, out-of-stock loss cost, transportation cost and delay time cost of manufacturers, suppliers and distributors.
- (2) Build the cold chain logistics inventory control dynamic model, and design the cold chain logistics inventory control algorithm.

## 2 Design of Inventory Control Method

### 2.1 Establish Inventory Cost Model of Cold Chain Logistics

How to improve the response speed of enterprises to market changes, restrain the bullwhip effect in inventory management, coordinate and optimize the resources of upstream and downstream enterprises in the supply chain, reduce inventory costs, and meet the personalized needs of customers through effective information technology have become the urgent problems to be solved in enterprise supply chain management. As an emerging disruptive information technology, the development of blockchain technology has completely changed the business model in many economic fields, especially in the field of logistics and supply chain management. By improving the accuracy, transparency and accessibility of information sharing in the supply chain, blockchain technology ensures trust among partners, less transaction cost, shorter delivery time and improved demand forecasting. Therefore, intuitively, new energy consumption technology focuses on improving the information quality of supply chain management, and changing the order fulfillment, distribution, intermediate product delivery, logistics and transportation in supply chain management [5].

The main problem of this paper is to design a multi echelon inventory model based on new energy consumption technology. Without changing the nature of the problem, we make the following assumptions for some complex problems.

First, assume that the information of supply chain is related to the application degree of blockchain technology, that is, the order quantity  $Q_i$  and demand quantity  $D_i$  of suppliers are positively related to the application degree  $\omega_i$  of new energy consumption technology, and  $0 \leq \omega_i \leq 1$ . The new energy consumption technology improves the accuracy of inventory information

$$D_i = Q_i + (1 - \omega_i^2)\zeta \tag{1}$$

In formula (1),  $D_i$  is the demand of the supplier;  $Q_i$  is the order quantity of the supplier;  $\omega_i$  is the correlation coefficient between demand and new energy consumption technology;  $\zeta$  is the random error.

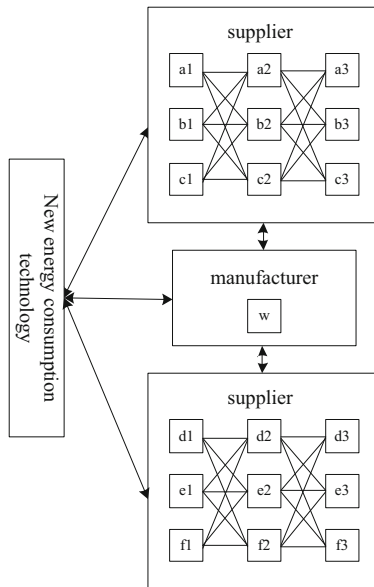
Second, inventory information quality problems lead to a longer delay time.

Third, the supply chain can produce a variety of core products, allowing a variety of raw materials, a supplier only provides one raw material.

Fourth, the daily demand of each node in the supply chain is random and follows the normal distribution [6].

Fifthly, all enterprises adopt the inventory management strategy of random demand and random lead time  $(\alpha, \beta)$ .

Sixth, the order lead time function is a random distribution function. Seventh, stochastic lead time is independent of each other, and lead time and demand are independent of each other. The flow direction of logistics between different levels is shown in Fig. 1.



**Fig. 1.** Inventory cost model under the influence of new energy consumption technology

As shown in Fig. 1, in this model, the connecting line represents the connection between nodes, while the double arrow indicates that there are both forward logistics

and reverse logistics between enterprises at different levels. In general, double arrows can also be converted to single arrows to indicate that high-level nodes can only transmit in one direction, but not from low-level nodes to high-level nodes [7].

### 2.2 Calculate the Inventory Cost of Cold Chain Logistics

After analyzing the causes of inventory cost, this model divides inventory cost into order cost, inventory holding cost, shortage loss cost, transportation cost and delay time cost. The formula for calculating the manufacturer’s annual total ordering cost is as follows:

$$Y_{zz}^{order} = \sum_{i=1}^{m_n} \sum_{j=1}^{m_{n-1}} \frac{X_n^{(d,g)} N_n^{(d,g)}}{d_n^{(d,g)}} \tag{2}$$

In Eq. (2),  $X_n^{(d,g)}$  is the order cost;  $N_n^{(d,g)}$  is the annual order quantity;  $d_n^{(d,g)}$  is the order quantity of the upper level enterprise to the lower level enterprise. The formula for calculating the annual total ordering cost of the supplier is as follows:

$$Y_{gy}^{order} = \sum_{i=1}^{m_{n-1}} \sum_{j=1}^{m_{n-2}} \sum_{v=1}^{m_{n-(n-2)}} \frac{X_{n-1}^{(d,g)} N_{n-1}^{(d,g)}}{d_{n-1}^{(d,g)}} \tag{3}$$

In formula (3),  $X_{n-1}^{(d,g)}$  represents the ordering cost of the supplier;  $N_{n-1}^{(d,g)}$  is the annual order quantity of the supplier;  $d_{n-1}^{(d,g)}$  represents the supply quantity of the upper level enterprise to the lower level enterprise [8]. The annual total order cost of the distributor is calculated as follows:

$$Y_{fx}^{order} = \sum_{i=1}^{m_{n+1}} \sum_{j=1}^{m_{n+2}} \sum_{v=1}^{m_{n+(n+2)}} \frac{X_{n+1}^{(d,g)} N_{n+1}^{(d,g)}}{d_{n+1}^{(d,g)}} \tag{4}$$

In Eq. (4),  $X_{n+1}^{(d,g)}$  is the order cost of the distributor;  $N_{n+1}^{(d,g)}$  is the annual order quantity of the distributor;  $d_{n+1}^{(d,g)}$  represents the sub sales volume of the upper level enterprise to the lower level enterprise. The total cost is calculated as follows:

$$Y_z^{order} = Y_{zz}^{order} + Y_{gy}^{order} + Y_{fx}^{order} \tag{5}$$

In Eq. (5),  $Y_z^{order}$  is the total order cost of the three-level logistics chain;  $Y_{zz}^{order}$  is the total annual ordering cost of the manufacturer;  $Y_{gy}^{order}$  is the total annual ordering cost of the supplier;  $Y_{fx}^{order}$  is the total annual ordering cost of the supplier [9]. In the calculation of inventory cost, the above formula is basically followed. The total inventory cost is:

$$Y_z^{hold} = Y_{zz}^{hold} + Y_{gy}^{hold} + Y_{fx}^{hold} \tag{6}$$

In Eq. (6),  $Y_z^{hold}$  represents the total inventory cost in the flow process of the three-level logistics chain;  $Y_{zz}^{hold}$  is the total inventory cost of the manufacturer;  $Y_{gy}^{hold}$  is the total inventory cost of the supplier;  $Y_{fx}^{hold}$  is the total inventory cost of the supplier. The formula for calculating the total inventory cost of a manufacturer is as follows:

$$Y_{zz}^{order} = \sum_{i=1}^{m_n} \sum_{j=1}^{m_{n-1}} \frac{d_n^{(d,g)}}{2 + R_n^{(d,g)} - W_n^{(d,g)}} \tag{7}$$

In Eq. (7),  $d_n^{(d,g)}$  represents the inventory cost in the supply process of the upper level enterprise to the lower level enterprise;  $R_n^{(d,g)}$  is the manufacturer’s reorder inventory;  $W_n^{(d,g)}$  is the safety stock of the supplier. The formula for calculating the annual total ordering cost of the supplier is as follows:

$$Y_{gy}^{order} = \sum_{i=1}^{m_{n-1}} \sum_{j=1}^{m_{n-2}} \sum_{v=1}^{m_{n-(n-2)}} \frac{d_{n-1}^{(d,g)}}{2 + R_{n-1}^{(d,g)} - W_{n-1}^{(d,g)}} \tag{8}$$

In Eq. (8),  $d_{n-1}^{(d,g)}$  represents the inventory cost of the upper level enterprise in the process of supplying goods to the lower level enterprise;  $R_{n-1}^{(d,g)}$  is the reorder inventory of the supplier;  $W_{n-1}^{(d,g)}$  is the safety stock of the supplier [10]. The annual total order cost of the distributor is calculated as follows:

$$Y_{fx}^{order} = \sum_{i=1}^{m_{n+1}} \sum_{j=1}^{m_{n+2}} \sum_{v=1}^{m_{n+(n+2)}} \frac{d_{n+1}^{(d,g)}}{2 + R_{n+1}^{(d,g)} - W_{n+1}^{(d,g)}} \tag{9}$$

In Eq. (9),  $d_{n+1}^{(d,g)}$  represents the inventory cost of the sales process of the upper level enterprise to the lower level enterprise;  $R_{n+1}^{(d,g)}$  is the reorder inventory of the distributor;  $W_{n+1}^{(d,g)}$  is the safety stock of the distributor [11]. By analogy, the shortage cost, transportation cost, loss and delivery delay cost are calculated respectively.

According to relevant research, new energy consumption technology can accurately share the demand, inventory and transportation information of the upstream and downstream node enterprises in the supply chain, provide high-precision and reliable prediction for warehousing and logistics activities, and reduce the delay time. In order to reflect the influence of the timeliness of inventory information of new energy consumption technology, it is assumed that the calculation formula of delay time is as follows:

$$T_{yc} = \chi_g \log_b^{\chi_g} \tag{10}$$

In formula (10),  $T_{yc}$  represents the total time of the delay of cold chain logistics;  $\chi_g$  is the delay parameter, usually taken as an integer between 100–300;  $b$ ; is the negative correlation index of delay time and new energy consumption technology, which is usually taken as 0.5.

### 2.3 Build the Dynamic Model of Cold Chain Logistics Inventory Control

Through Markov chain and Markov decision-making process, this paper establishes four tuples of reinforcement learning (environmental state observation, agent action, state transfer, reward), and applies reinforcement learning algorithm to fresh inventory control model. The following mainly describes the relationship between state variables, action variables, reward and punishment function, product life cycle, lead time and decay rate [12]. The state of product directly affects the decision of agent. In this paper, the state variable is set as  $D_y$ , and the relationship between the remaining life cycle  $L_s$ , life cycle  $L_s$  and the retailer's remaining inventory  $N_l$  is as follows

$$D_y(s) = [L_s, N_l] \quad (11)$$

In Eq. (11),  $D_y(s)$  is the correlation index between the remaining life cycle  $L_s$  of cold chain inventory products and the remaining inventory  $N_l$  of retailer products in the case of S. In this process, if the product quantity is insufficient, it will be regarded as out of stock, if the product inventory is sufficient. However, if the product is not in the life cycle, it is considered to be out of date. If the product is in the life cycle and has sufficient inventory, it is considered to be saleable [13]. State  $D_y$  is related to inventory quantity  $N_l$ , so  $0 \leq D_y \leq 100$  is satisfied. In order to make the objective function converge faster, the continuous state can be discretized and divided into 100 integers in the [0100] interval. In order to fit the actual business state better and discretize the objective function, we can get the following results:

$$B_i = [d_i] \quad (12)$$

In Eq. (12),  $B_i$  represents the agent's action variable, usually  $B_i \in [0, 10d_i]$ ;  $d_i$  represents the action of the agent, and the value is [0, 10], which is generally a positive integer [14]. The action variable is usually a decision-making item after observation. After integrating the reward and punishment variables, the following formula can be obtained:

$$f_i(x) = C_g \times l_g + C_q \times l_q \quad (13)$$

In Eq. (13),  $f_i(x)$  represents the reward and punishment relationship of the algorithm;  $C_g$  is the cost of expired materials in cold chain logistics inventory;  $l_g$  is the quantity of expired materials in cold chain logistics inventory;  $C_q$  is the cost of scarce materials in cold chain logistics inventory;  $l_q$  is the quantity of scarce materials in cold chain logistics inventory.

The life cycle of a product is the only standard to measure whether a product is overdue. The shorter the life cycle of cold chain products is, the more strict the retailer's order strategy is. In view of the fact that the best shelf life of most cold chain products on the market is 3 days, this paper defines the life cycle of products as 3 days and the lead time of products as 1 day. The remaining life of a product is the difference between the product life cycle and the lead time.

$$L_i = L_T - M_T \quad (14)$$

In Eq. (14),  $L_i$  is the remaining life of cold chain logistics products;  $L_T$  is the average life cycle of cold chain logistics products;  $M_T$  is the lead time of cold chain products [15]. In addition, we also need to calculate the damage rate of products (damage rate), which represents the probability of unit quantity of cold chain logistics products in the process of transportation and storage. Since most of the cold chain products decay occurs in the storage stage, this paper defines the decay rate  $D_{sfl}$  as the quotient of  $f_i(x)$  and the total product value  $V_Z$ . The formula is as follows:

$$D_{sfl} = \frac{f_i(x)}{V_Z} \quad (15)$$

In Eq. (15),  $D_{sfl}$  is the damage rate of cold chain logistics products;  $f_i(x)$  is the relationship between reward and punishment;  $V_Z$  is the total value of cold chain logistics products.

#### 2.4 Design Cold Chain Logistics Inventory Control Algorithm

The optimization model studied in this chapter belongs to multi constraint single objective nonlinear problem, involving multiple decision variables, and the solving process is very complex. Therefore, it is necessary to design a random search algorithm based on biological natural selection and natural genetic mechanism, which has strong robustness and global search efficiency for solving multi-objective optimization problems, and can overcome the fast decline trap problem of other algorithms, so that it does not fall into the local optimum. When the above algorithm is applied to solve the problem, the initial enterprise level coding is carried out first, and different enterprise level coding needs different model problems, which also affects the operation steps of the algorithm. This paper gives a strict explanation of enterprise coding and points out that enterprise level coding is the mapping from phenotype of spatial data to genotype data. The commonly used enterprise level codes are binary code and real code [16]. There are practical problems in binary coding, which can affect the coding efficiency of the algorithm, resulting in a large distance between binary codes of adjacent integers. Especially in the face of more variables, more complex problems and a wide range of constraints, the convergence speed of genetic algorithm is significantly inhibited. In view of this, real coding has the advantage of solving complex multivariable optimization problems, and has high accuracy and stability, and can accelerate the convergence speed of the algorithm. Based on the above concepts, this paper designs the model algorithm as shown in Fig. 2.

Through small probability adjustment, one or some bit values on the individual coding string can be changed [17]. The operation of this algorithm is to replace the gene values of some loci in the enterprise coding string with other alleles of the locus, so as to form a new individual. It can avoid some information loss caused by selection and crossover operation, and ensure the effectiveness of the algorithm. The crossover operator determines the local search ability of the algorithm, and the operation is an auxiliary method to generate new individuals. Through the above algorithm, we can

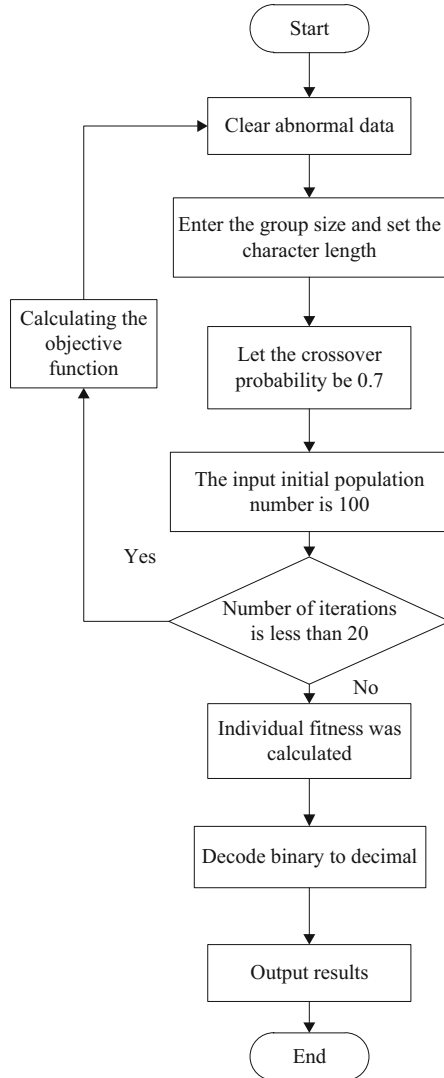


Fig. 2. Model algorithm design

realize the optimization of inventory control of cold chain logistics enterprises under the background of new energy consumption.

### 3 Experimental Study

In order to verify the effectiveness of inventory control method of cold chain logistics enterprises under the background of new energy consumption, the following experiments are designed.



### 3.1 Experimental Design and Preparation

This experiment mainly uses frequency response to measure bullwhip effect in supply chain, so as to verify the inhibitory effect of variables introduced in two-level inventory on information delay and bullwhip effect in Multi-level Inventory, so as to prove the effectiveness, rationality and innovation of the inventory control method in this paper.

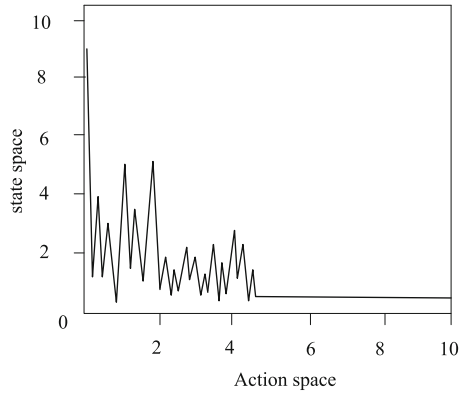
The frequency characteristic of the control system reflects the response performance of the system under the action of sinusoidal signal, and the frequency domain analysis method is to use the frequency characteristic to study the linear system. Taking the minimum inventory cost of cold chain logistics enterprises as the goal, this paper constructs a single product inventory control model with single-stage supply chain as the core through the above research. Considering the actual factors such as damage rate, overdue cost, shortage cost and so on, combined with the best learning rate and exploration rate, the state space and action space of the inventory control method designed in this paper are scaled up, At the same time, the user demand data based on Poisson distribution is enlarged according to the corresponding proportion. Through the simulation experiment, the product damage rate of the three traditional inventory control methods is compared. In order to get the best results, the state, action and experimental data of the two algorithms are defined as follows.

The inventory control method designed in this paper is taken as the experimental group. In the model,  $E \in [0, 100]$  is taken as a positive integer. Due to the limitations of the action space and state space of the inventory control method in the experimental group, the values of action space and state space are  $A \in [0, 10B]$  and  $B \in [0, 10]$  respectively. In the experiment, three groups of customer demand data based on Poisson distribution with different eigenvalues are set through the characteristics of inventory capacity, which are the distribution rule based on parameter  $\lambda(10, 20, 30)$ . Each group of data is 1000.

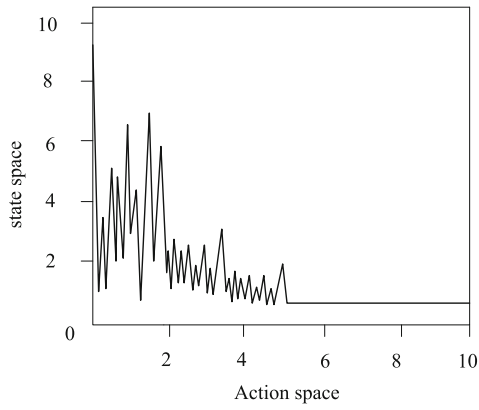
Three traditional methods are used as three control groups. In control group 1,  $E_1 \in [0, 100]$  is a positive integer, and the values of action space and state space are  $A_1 \in [0, 20B]$  and  $B_1 \in [0, 25]$  respectively. In control group 2, the inventory control method  $E_2 \in [0, 200]$  was a positive integer, and the action space and state space were  $A_2 \in [0, 25B]$  and  $B_2 \in [0, 40]$ , respectively. In control group 3, the inventory control method  $E_3 \in [0, 200]$  was a positive integer, and the values of action space and state space were  $A_3 \in [0, 25B]$  and  $B_3 \in [0, 50]$ , respectively. According to the inventory capacity characteristics of three control groups, three groups of customer demand data based on Poisson distribution with different eigenvalues are set in the experiment, which are user demand data based on parameter  $\lambda_1(20, 40, 60)$ ,  $\lambda_2(100, 200, 300)$ ,  $\lambda_3(40, 80, 120)$  distribution, with 1000 data in each group.

### 3.2 Model Performance Test

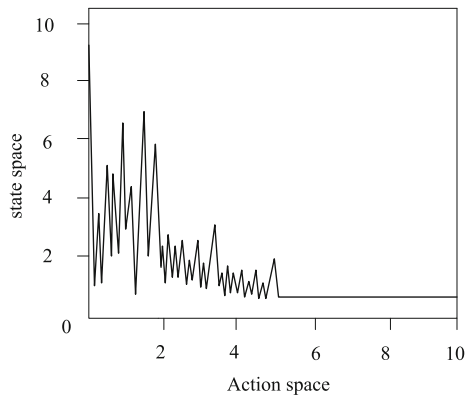
The model of the experimental group and the three control groups were tested under different parameters, and the results are shown in Fig. 3.



(a) Experience group

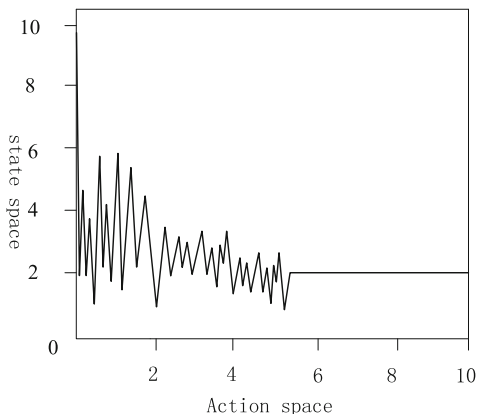


(b) Control group 1



(c) Control group 2

**Fig. 3.** Experimental performance test



(d) Control group 3

**Fig. 3.** (continued)

As shown in Fig. 3, the experimental results shown in Table 1 can be obtained from the above test results.

**Table 1.** Results under different parameters

Group	Parameter	Mean value of results
Experience group	10	2.3652
	20	2.3524
	30	2.0125
Control group 1	20	1.2365
	40	1.6354
	60	1.9635
Control group 2	100	1.3256
	200	1.3642
	300	1.5246
Control group 3	40	1.7586
	80	1.9635
	120	1.8515

In Table 1, through the experimental results under different parameters, we can know that the performance of the inventory control method designed in this paper is better than the three traditional inventory control methods, and its ability to process data is stronger. In this method, when the action space is about 4.658, the stability of the state space can be achieved, so the decision stability of this method is higher than that of the other three methods.

### 3.3 Effect Test of Inventory Control Method

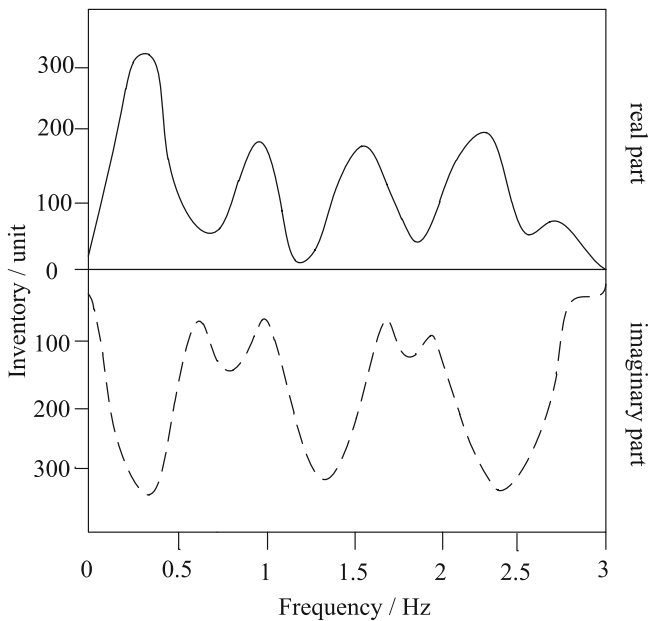
In this experiment, we assume that any customer needs to be composed of multiple sinusoidal waves with different frequencies and amplitudes. Therefore, the frequency response amplitude  $|F(\omega)|$  is calculated, and the frequency response curves of order quantity and inventory quantity under different parameters are described by MATLAB programming software.

$$|F(\omega)|=|G(\zeta)|=\sqrt{X(G(\zeta))+ln^2(G(\zeta))} \quad (16)$$

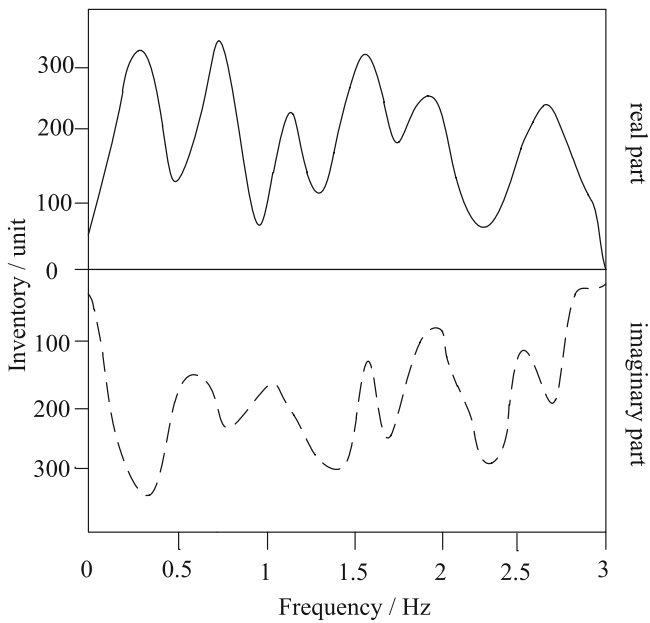
In Eq. (16),  $X(G(\zeta))$  represents the value of the real part of the frequency domain;  $ln^2(G(\zeta))$  is the value of the imaginary part of the frequency domain. Synthesize the experimental data obtained by formula (16), and draw the final effect test curve under four inventory control methods as shown in Fig. 4.

In this section, Matlab is used for computer simulation analysis of the above model, and the amplitude of dynamic response of inventory under different demand is compared, as shown in Table 2.

According to the data in Table 2 and Fig. 4, combined with the measurement principle of bullwhip effect, we can get the following conclusions: firstly, from the overall data, the parameter variable is the best  $\mu$ . The smaller the inhibition was, the more significant the inhibition was  $\mu =$  The average threshold is about 0.22 at 1  $\mu = 2$ , the average threshold is about 0.33  $\mu =$  The average threshold is about 0.37. Secondly, the effect test is shown in Fig. 7  $\mu$ . In the case of fixed demand, variables have better inhibitory effect on order variability than in the case of random fixed demand. In view of the problem of product information delay in inventory, it can be concluded from Table 2 that the greater the information delay, the more the order fluctuation. And  $\mu$  Variables can effectively suppress the bullwhip effect caused by information delay. In addition, even if the information delay time is 0, there will be order fluctuation or inventory amplification problem. When the random demand signal is input, the longer the information delay time in the supply chain system, the exponential smooth development, which is extremely effective to suppress the order amplification effect of the supply chain. Finally, the above analysis verifies the effectiveness and superiority of the inventory control method of cold chain logistics enterprises designed in this paper.

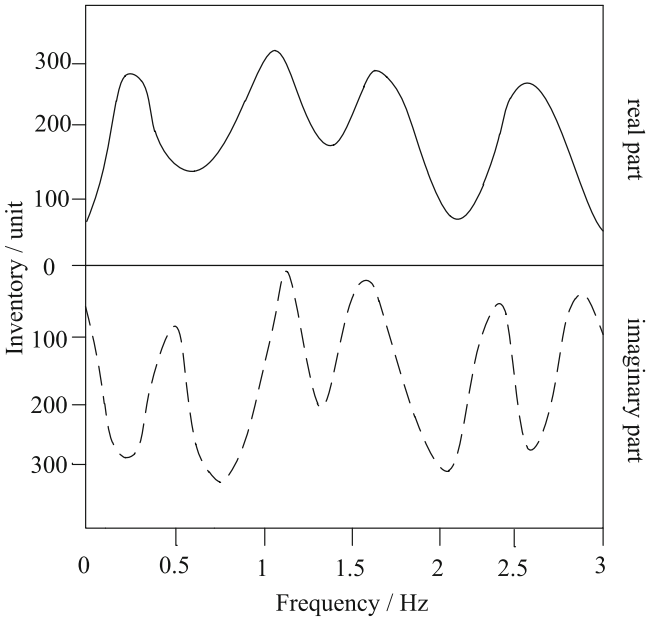


(a) Experience group

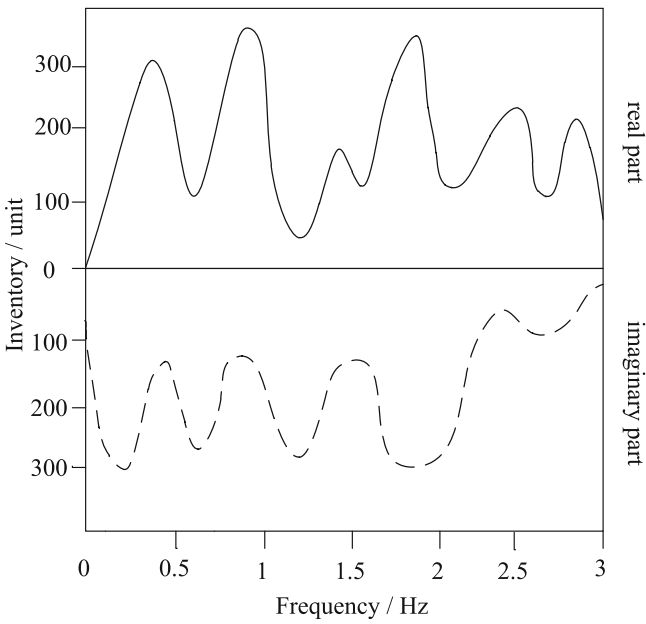


(b) Control group 1

Fig. 4. Experimental effect test



(c) Control group 2



(d) Control group 3

**Fig. 4.** (continued)

**Table 2.** Inventory dynamic response threshold

Group	Frequency domain	$\mu = 1$	$\mu = 2$	$\mu = 3$
Experience group	Real part	0.25	0.35	0.34
	Imaginary part	0.25	0.36	0.36
Control group 1	Real part	0.25	0.34	0.39
	Imaginary part	0.26	0.32	0.38
Control group 2	Real part	0.24	0.31	0.38
	Imaginary part	0.21	0.30	0.37
Control group 3	Real part	0.14	0.34	0.37
	Imaginary part	0.16	0.33	0.36

## 4 Conclusion

The internal competition of the supply chain system is diversified and complex, which aggravates the information asymmetry between the supply chain nodes and increases the inventory backlog and management difficulty. As an important breakthrough of independent innovation of China's core technology, new energy consumption technology can effectively solve the increasingly severe problems of low data transparency, difficult traceability, lack of trust mechanism, information island and so on in supply chain management. On the basis of exploring the influence of new energy consumption technology on the upstream and downstream enterprises of supply chain, this paper studies the static and dynamic models of Multi-level Inventory of supply chain by using mathematical model and cybernetics, analyzes the optimization of complex supply chain system under the background of new era, and enriches the research of new energy consumption technology in supply chain management. Experiments are designed to verify the reliability and superiority of the proposed method.

## References

1. Wang, L.: Research on curriculum teaching educational reform based on DCCPP cultivation mode—taking inventory control and management as an example. *Logist. Eng. Manag.* **43**(2), 149–152 (2021)
2. Sun, P., Sun, R., Li, J.: Inventory cost control model of fresh product retailers based on reinforcement learning. *Comput. Simul.* **37**(8), 192–195+201 (2020)
3. Sun, D., Song, J., Zhang, Y.: Simulation of inventory control model for emergency grain logistics network after disaster. *Comput. Simul.* **37**(3), 470–474 (2020)
4. Ye, B., Chen, Y., Zhang, B., et al.: Research on multi-echelon coordinated inventory control model of tobacco industry enterprises. *Acta Tabacaria Sinica* **26**(4), 117–123 (2020)
5. Wang, F., Qi, Y., Long, J., et al.: Optimal inventory control under logistics path closure and congestion risks. *J. Syst. Manag.* **29**(5), 964–973 (2020)
6. Sheng, Q., Zhang, J., Sun, J.: Research on inventory control strategy of a manufacturing enterprise. *Mach. Des. Manuf.* **45**(11), 301–304 (2020)

7. Hong, F., Wang, L., Zheng, J.: Category management and inventory control strategy for multi-regional network retail with alternative demand. *Logist. Technol.* **38**(11), 51–57 (2019)
8. Chen, W., Zhang, S., Hu, X., et al.: Inventory control of fresh food supply chain and its differential game considering fresh-keeping effort. *J. Univ. Sci. Technol. China* **49**(6), 494–505 (2019)
9. Xie, R., Zhao, M.: Research on enterprise inventory control based on RBF neural network technology. *Comput. Digit. Eng.* **47**(12), 3078–3081 (2019)
10. Lou, S., Rong, X.: Inventory control under demands and returns dependent supply disruptions. *Control Decis.* **36**(4), 1003–1009 (2021)
11. Chen, H., Zhang, L.: Application of activity based classification and purchase restriction method in controlling drug inventory in our hospital. *Chin. J. Med. Guide* **22**(3), 197–200 (2020)
12. Zhang, Y.: Research on supply chain management inventory control based on price discount: taking G online pharmaceutical company as an example. *Value Eng.* **40**(5), 100–101 (2021)
13. Di, J., Gong, H., Sun, Z.: Optimal control strategy of service-inventory system with Erlang distributed lead time. *J. Changsha Univ. Sci. Technol. (Nat. Sci.)* **16**(3), 90–98 (2019)
14. Wang, S., Yue, D., Zhang, Y., et al.: Optimal control of M/M/1 production inventory system with positive service time and perishable items. *Chin. J. Eng. Math.* **37**(6), 685–698 (2020)
15. Chen, Z., Zeng, F., Li, J., et al.: Inventory control management of medical materials based on ABC-VED matrix analysis. *China Med. Devices* **34**(10), 155–157 (2019)
16. Zou, Y., Liu, Q., Ye, C., et al.: Optimization research of inventory-production strategy for dairy cold chain logistics based on quality level. *Appl. Res. Comput.* **37**(9), 2744–2748 +2753 (2020)
17. Zhou, B., Zong, S.: Lagrangian relaxation algorithm for cross-dock scheduling & problems in cold-chain logistics. *Control Theory Appl.* **37**(3), 505–512 (2020)





# Fault Detection Method of Electronic Circuit Converter Based on Dynamic Sequence Response

Dan-kang He<sup>(✉)</sup> and Qiu-jiao Huang

Guangxi Modern Polytechnic College, Hechi 547000, China  
liangxiaodan7788@163.com

**Abstract.** Traditional electronic circuit converter fault detection methods take local eigenvalues as eigenvectors to analyze. In the actual work process, different factors will lead to the difference of electronic circuit converter features, which leads to the low accuracy of feature classification. Therefore, this paper designs an electronic circuit converter fault detection method based on dynamic sequence response. Firstly, the data collector is designed to complete the collection of electronic circuit converter signal, the analysis model based on dynamic sequence response analysis method is established, the short-circuit fault signal feature extraction based on wavelet transform, the fault detection classifier is optimized, the fault location algorithm is optimized, and the design of electronic circuit converter fault detection method based on dynamic sequence response is completed. The experimental results show that the accuracy of the traditional method is only 25.3%, and the accuracy of the designed method can reach 88.7%, which verifies the effectiveness of the design method.

**Keywords:** Dynamic sequence response · Electronic circuit · Converter fault detection

## 1 Introduction

With the rapid development of power electronic technology and the continuous expansion of power electronic device market, as the core part of the whole power conversion, electronic circuit converter industry, energy, transportation, national defense and other fields play an increasingly important role. According to statistics, about 75% of electric energy in developed countries is used after power electronic converter technology, which is expected to reach more than 95% soon. However, the following converter circuit fault problem is also increasingly prominent [1–3]. Due to the wide use and importance of power electronic converter technology in various application fields, the reliability of these devices with converter circuits is required to be higher and higher. Once the fault occurs, especially the open circuit fault of the device, if the fault of the converter circuit can not be correctly diagnosed and repaired in time, it will cause waveform distortion and system instability, or direct shutdown, or even damage the equipment and personnel, and bring huge economic losses and accidents. Especially in aviation, national defense and other application fields with high

requirements for equipment reliability, it is more necessary to ensure the timely and accurate diagnosis of converter circuit device in case of fault [4, 5]. Therefore, it is of great significance to propose intelligent, fast and accurate fault diagnosis method for power electronic converter circuit.

In reference [6], an improved floating board modulator and a fault detection circuit for positive and negative bias voltage are proposed. By using the parasitic capacitance characteristics of MOSFET, the modulation pulse width is controlled by a fixed pulse width, and the output negative bias voltage is increased by adding a negative bias MOSFET. In reference [7], a fault protection scheme for overhead flexible direct power grid with reactor voltage difference is proposed. In this scheme, the difference of positive and negative reactor voltage amplitude is used for fault pole selection, and the sum of positive and negative reactor voltage is used for fault detection.

In the traditional detection methods of electronic circuit converter fault, the current time and current value and other local eigenvalues are used as eigenvectors to analyze. However, in the actual working process, the operating conditions, manufacturing process, circuit breaker model and operating time will lead to certain differences in the current eigenvalues of electronic circuits. The traditional methods have certain limitations. The accuracy of feature classification is low. Therefore, this paper designs an electronic circuit converter fault detection method based on dynamic sequence response. The data collector is designed to complete the collection of electronic circuit converter signal, the analysis model based on dynamic sequence response analysis method is established, the feature extraction of short circuit fault signal based on wavelet transform, the fault detection classifier is optimized, the fault location algorithm is optimized, and the electronic circuit converter fault detection based on dynamic sequence response is completed.

## **2 Fault Detection Method of Electronic Circuit Converter Based on Dynamic Sequence Response**

### **2.1 Acquisition of Electronic Circuit Converter Signal**

For the current signal acquisition of electronic circuit, the most commonly used method is to use the open-ended Hall current sensor. In the measurement process, the incoming line on one side of the electromagnet in the electronic circuit is connected with the sensor, and the acquisition signal is transmitted from the sensor to the data acquisition card through the NBC interface [8–10]. In the process of current signal measurement, the collected electronic circuit converter signal will be affected by the electromagnetic interference in the process of equipment operation, external environment and transmission, including some random and irregular noise, which makes the current signal distorted. These noises will disturb the characteristic information of the original current waveform.

Data collector is an important source of monitoring electronic circuit operation data, which mainly completes the voltage and current phasor acquisition. FPGA is mainly used to control the ad chip. After receiving the GPS clock signal, FPGA triggers the ad chip through a series of processing such as frequency division [11–13]. In order to ensure the requirements of data acquisition and transmission, a SMV (sampling measurement value) interface is designed in the data collector. The interface can efficiently and quickly complete the acquisition and transmission of waveform data, and reduce the delay. Using SMV for fast transmission of sampling points can solve the throughput capacity of big data. The data transmission of SMV message occurs at the data link layer, and its message format is shown in the following table (Table 1):

**Table 1.** SMV message format

MAC frame content	Explain	Number of bytes
Preamble (preamble)	The receiver realizes synchronization and extracts clock information	8 bytes
Frame start delimiter (SFD)	Represents the beginning of a frame. 1 and 0 in the field are used interactively to mark the beginning of valid data	1 byte
Header MAC	Mac destination address	16 bytes
Priority tagged	$0 \times 210XXA03000-0 \times 210XXA0303ZZ$	2 bytes
Byte padding	It is used to distinguish the eight priorities	0–32 bytes
CRC check	$TPIO = 0 \times 2100$	4 bytes
Ethertype	$APPID = 0 \times 2000-0 \times 210XX$	2 bytes

Eight analog channels are designed in the collector, each channel is compatible with current and voltage, and the input current range is 0–5 A, the input voltage range is 0–350 V.

The data collector of electronic circuit needs to have the function of phasor measurement in the process of operation. It needs to have a higher sampling range, and make the data signal of the acquisition band present the completed waveform. Therefore, the data collector needs to have the characteristics of high range and high precision, and it needs to add current transformer and voltage transformer [14–16]. TS24D43-100 A/30 mA is used as the current transformer. The measurement range of this type of transformer is wide and the accuracy is high. The voltage transformer is ZWQ281A. The conversion accuracy of this type of transformer is very high, which is suitable for the monitoring of micro equipment. The detailed parameters of the two transformers are shown in the table below (Table 2):

**Table 2.** Transformer parameters

Transformer type	Project	Detailed parameters
Current transformer	Rated input/output current (mA)	10 A/0.5
	Rated point difference (%)	$\leq \pm 0.1$
	Load ( $\Omega$ )	$\leq 120$
	Accuracy class	0.1
	Linear range	5%In–2000%In
Voltage transformer	Rated input current (mA)	1 A
	Rated output current (mA)	1 A
	Transformation ratio	1000: 1000
	Phase difference	$\leq 20'$ (input is 1a, sampling resistance is 100 $\Omega$ )
	Linear range	0–1000 V, (sampling resistance is 100 $\Omega$ )
	Permissible error	$-5\% \leq f \leq 0$ (sampling resistance is 100 $\Omega$ )
	Working temperature	$-40\text{ }^\circ\text{C} \pm 85\text{ }^\circ\text{C}$

After the data collected is processed by the two transformers, the sampling value sequence is obtained, which has 8 channels, each channel has 8 bytes, which represents the effective data quality of the sampling value, which determines the effectiveness of the channel [17, 18].

In order to detect the current change fault of electronic circuit, it is necessary to extract the noise free waveform of the electronic circuit, so as to improve the accuracy of fault detection. Therefore, the collected electronic circuit is transformed into a stream. In this paper, the five point three smoothing method is used to filter the noise. Assuming the format of the experimental data is  $Y_{-n}, Y_{-n+1}, \dots, Y_0, Y_1, \dots, Y_n$ , some isometric nodes  $X_{-n}, X_{-n+1}, \dots, X_0, X_1, \dots, X_n$  are set, and the distance between the equidistant nodes is  $h$ . The nesting between the experimental data and the equidistant nodes can be completed. After the relevant exchange calculation, the data can be nested with the equidistant nodes. The experimental data can be fitted and calculated

$$Y(t) = a_0 + a_1t + a_2t^2 + \dots + a_mt^m \tag{1}$$

In the above formula,  $t$  can be expressed as  $t = \frac{x-x_0}{h}$ , which is the functional form of exchange calculation. For the undetermined coefficients in the above formula, the least square method is needed to solve them:

$$\sum_{i=-n}^n R_i^2 = \sum_{i=-n}^n \left[ \sum_{j=0}^m a_j t_i^j - Y_i \right]^2 = \phi(a_0, a_1, \dots, a_m) \tag{2}$$

For the above formula, the objective of the least square method is to minimize  $\phi(a_0, a_1, \dots, a_m)$ . Therefore, it is necessary to find the partial derivative of the above formula for  $a_k (k = 0, 1, \dots, m)$ , so that the partial derivative is 0, and the derivative equation of the following formula is obtained:

$$\sum_{i=-n}^n Y_j t_i^k = \sum_{j=0}^m a_j \sum_{i=-n}^n t_i^{k+1} \tag{3}$$

In this paper, the five point cubic smoothing method is used to carry out simple noise filtering, that is, the values of  $n, m$  in the above formula are 2 and 3 respectively, so that five nodes can be obtained, and  $t$  can be assigned to obtain the final five point cubic smoothing equation system

$$\begin{cases} \bar{Y}_{-2} = \frac{1}{70}(69Y_{-2} + 4Y_{-1} - 6Y_0 + 4Y_1 - Y_2) \\ \bar{Y}_{-1} = \frac{1}{30}(2Y_{-2} + 27Y_{-1} + 12Y_0 - 8Y_1 + 2Y_2) \\ \bar{Y}_0 = \frac{1}{35}(-3Y_{-2} + 12Y_{-1} + 18Y_0 + 12Y_1 - 3Y_2) \\ \bar{Y}_1 = \frac{1}{35}(2Y_{-2} - 8Y_{-1} + 12Y_0 + 27Y_1 + 2Y_2) \\ \bar{Y}_2 = \frac{1}{70}(-Y_{-2} + 4Y_{-1} - 6Y_0 + 4Y_1 + 69Y_2) \end{cases} \tag{4}$$

The de-noising method selected in this paper requires at least five nodes. In order to ensure the symmetry of the original electronic circuit converter signal in the de-noising process, different equations need to be selected for processing in different parts of the electronic circuit converter signal [19–21].

## 2.2 The Analysis Model Based on Dynamic Sequence Response Analysis Method is Established

Dynamic sequence response analysis is to arrange a group of random sequences in order of time sequence, and use statistical method to deal with the dynamic sequence response value. In statistics, the less variables are analyzed, the simpler the process is, the higher the accuracy of the results is. Firstly, the historical characteristic data of the electronic circuit to be tested is obtained, and it is taken as a sample set  $x$ , and the samples are respectively expressed as  $x_1, x_2, \dots, x_n$ . Then the maximum value, minimum value, average value and standard deviation of the sample set can be calculated, and the curve fitting can be performed to fit the distribution according to the probability distribution of the variable flow characteristic data. The distribution curve is tested by the method of distribution fitting goodness test. The chi square test is used. The formula of chi square value is as follows:

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - m_i)^2}{m_i} = \sum_{i=1}^r \frac{(n_i - Np_i)^2}{Np_i} \tag{5}$$

In the above formula,  $n_i$  is the measured frequency of group  $i$ ,  $m_i$  is the theoretical frequency,  $p_i$  is the probability that the observation results fall into group  $i$ , and  $N$  is the

total number of samples. In the dynamic sequence response analysis algorithm, the correlation quantization calculation needs to pay attention to the current conversion data of the adjacent points of the observation point:

$$x_{t,w} = [x_t, x_{t+1}, \dots, x_{t+w-1}]^T \quad (6)$$

In the above formula,  $t$  represents the initial time,  $w$  represents the sampling interval length, that is, the size of the window, and  $x_{t,w}$  represents the window at time  $t$ . The function of time window is to obtain the local covariance matrix of each dynamic series response. In the selection process of time window, the weighted exponential sliding time window is selected as the basis of dynamic series response analysis. The interval of the time window is  $1 \leq \tau \leq t$ , and each window is multiplied by a factor  $\beta^{t-\tau}$  to measure the value close to  $t$ , so the higher the weight is, the closer it is to the time window at  $t$ . When a dynamic sequence response  $T$  is known, the estimation formula of local autocovariance matrix at  $t$  time based on weighted exponential sliding time window is as follows:

$$\hat{\Gamma}_t = \hat{\Gamma}_{t-1} - T_{t-w,w} \otimes T_{t-w,w} + T_{t-w} \otimes T_{t-w} \quad (7)$$

In the above formula,  $\hat{\Gamma}_t$  represents the estimation result of the local autocorrelation matrix. It can be seen from the above formula that the estimation formula has nothing to do with the  $w$  values in the earlier stage, which can improve the calculation efficiency. Because the matrix of similarity value after decomposition has many components, it needs to be filtered before it can be calculated and used [22, 23].

### 2.3 Feature Extraction of Short Circuit Fault Signal Based on Wavelet Transform

On the basis of the above analysis model based on dynamic sequence response analysis method, in order to realize the current transformation fault detection of electronic circuit, this paper uses wavelet transform to extract the short circuit fault signal of sub circuit to realize the fault detection. It is necessary to analyze the fault signal accurately to locate the fault quickly for quick maintenance. By establishing the identification model and comparing the amplitude and frequency components of the fault signal in the electronic circuit information database, the existing fault symptoms can be judged and analyzed quickly and accurately. Therefore, the complex wavelet transform is helpful to improve the fault diagnosis rate.

In the feature extraction of electronic circuit converter fault, the wavelet is the moving local signal in the fault:

$$D = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (8)$$

Where  $\varepsilon^i(e)$  is the wavelet basis and  $s$  is the integrable function.

The wavelet function can accurately determine the fluctuation in the process of fault signal alternation, make it controlled in a reasonable range, and make it 0 in the average fluctuation range:

$$\int_{-\infty}^{\infty} |\varepsilon^i(e)|^2 dt = 0 \quad (9)$$

In the process of electronic circuit fault feature extraction, wavelet transform selection is also very important. The physical quantities and values of different nodes are also different. The input values of fault data are standardized in the range of [0, 0.1] and [0.9, 1.0], and the curve changes very flat in this range. The function formula of curve normalization is as follows:

$$T_i = Z_i - \frac{Z_{\min}}{Z_{\max}} - Z_{\min}\beta + \xi \quad (10)$$

Where,  $Z_i$  and  $T_i$  are the variables of normalization coefficient after wavelet transform,  $Z_{\min}$  is the minimum ordinate of the curve after wavelet transform,  $Z_{\max}$  is the maximum ordinate of the curve after wavelet transform,  $\beta$  is a constant in the process of curve transform, and the value of the constant will change according to the change of fault data input value, ranging from 1 to 1.5, take  $\xi = \frac{1-\beta}{2}$ . In the process of electromechanical fault detection, the nature and location of the fault are determined, including multi band energy, the time and amplitude of the maximum four points in these bands. These features are used to describe the input signal and serve as input signals. According to the travel time of wavelet in detection, the fault location is detected to realize the feature extraction of electronic circuit converter fault.

## 2.4 Fault Detection

Due to the wide distribution range of distribution network and numerous distribution equipment, the power consumption of distribution network devices is very different. Some old distribution lines and buildings are close to each other, and they are crossed or parallel with optical cables and various pipelines. When there is a fault, it will increase the complexity of troubleshooting. So we need to use fault location algorithm to detect fault information faster. At present, the original matrix algorithm can not determine the fault point at the end of the feeder, so the location algorithm is optimized.

The distribution network matrix of order  $M \times M$  and the problem cause matrix of order  $M \times M$  are constructed:

$$D = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{11}$$

Assuming that the number of distribution network nodes is  $M$ , when there is a feeder line between node  $i$  and node  $j$ , then the parameters in  $j$  column of each row  $i$  and  $i$  column of each row  $j$  in the matrix are represented by parameter 1. When there is information exceeding the maximum load current in FTU, a new fault information matrix  $G$  will appear

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{12}$$

When the fault current passes through the node  $j$  in the matrix  $G$ , that is, set the element  $G_{ij}$  to 1 and the others to 0. Multiply the matrix  $D$  and the matrix  $G$  to obtain the optimized fault location matrix  $P$ . When there is only one element in the device, in the event of a fault, after the fault current passes through the node  $i$ , the FTU of the node will transmit the information to the master station. After the master station receives the signal, the parameter of the matrix  $D$  is  $D_{ii} = 1$ , and the other elements in the matrix are set to 0. The signals received by the master station converge into an information collection matrix. When the parameter in the matrix is 1, it means that the  $i$  node has fault current and the direction is the same as the predicted direction; when the parameter in the matrix is 0, it means the  $i$  node. There is no fault current at each node; when the parameter in the matrix is  $-1$ , it means that there is an over fault current at the  $i$  node, but the direction is opposite to the predicted direction, so that the fault location can be completed.



### 3 Experiment

#### 3.1 Experimental Data

In order to verify the effectiveness of the coil current fault automatic diagnosis method based on artificial intelligence algorithm designed in this paper, the mechanical fault simulation of high voltage circuit breaker is carried out in the experiment. Five states are selected: buffer spring fatigue, low voltage of control circuit, closing spring fatigue, transmission looseness and normal state, These five states are simulated in the machine for testing. Due to the space, this paper lists the coil current characteristic parameters under three mechanical states, as shown in the table below (Table 3):

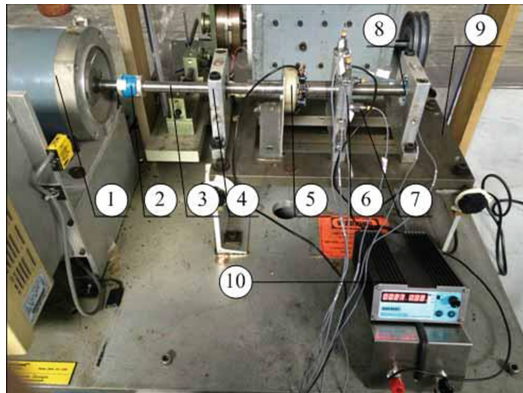
**Table 3.** Coil current characteristics under three mechanical conditions

Characteristic quantity	Fatigue of buffer spring	Low voltage of control circuit	Closing spring fatigue
$T_1/s$	0.022	0.056	0.034
$T_2/s$	0.036	0.059	0.043
$T_3/s$	0.045	0.078	0.057
$T_4/s$	0.091	0.094	0.073
$I_1/s$	0.95	1.31	1.03
$I_2/s$	0.54	0.73	0.85
$I_3/s$	0.91	1.15	1.38
$\mu$	0.154	0.256	0.354
$\sigma$	0.318	0.345	0.412
$K$	3.455	2.564	3.240
$W/J$	11.561	14.451	12.338

According to the characteristics of coil current in three different mechanical states in the above table,  $\mu$  represents mean value,  $\sigma$  represents standard deviation,  $K$  represents kurtosis, and  $W$  represents energy parameter. Under the above experimental conditions, the simulation platform selected in this paper is the Simulink board contained in MATLAB7.0, and the power components in the power system Simulation toolbox are used to build the internal circuit module of the electromechanical system of the expressway.

In the above experimental environment, the fault setting of the simulation circuit diagram is carried out. Taking the fault location, fault type and the change of grounding resistance as the simulation conditions, the denoising effect of the improved single ended A-type traveling wave location method and the improved semi soft threshold denoising method is comprehensively simulated. The operation status of the circuit breaker is evaluated by using the above current characteristic parameters. The automatic diagnosis method of coil current can be transformed into the problem of data classification. Using artificial intelligence algorithm to identify and classify the current characteristics under the same state, the coil current fault is identified.

According to the cause and mechanism of coil current fault, this paper designs an experimental device to simulate coil current, as shown in the figure below (Fig. 1):



**Fig. 1.** Experimental platform of electronic circuit converter damage

In the figure above, ① represents the motor, ② represents the insulated coupling, ③ represents the coil spindle, ④ represents the coil support seat, ⑤ represents the coil current loading device, ⑥ represents the support seat of the test coil, ⑦ represents the vibration acceleration sensor, ⑧ represents the insulated coil, and ⑨ represents the base. ⑩ Represents the coil current simulator. In the process of signal acquisition, this paper uses the data acquisition system of pilse, which is mainly composed of data acquisition card, vibration acceleration sensor and acquisition system.

Under the above experimental conditions, we use the method designed in this paper and the traditional electronic circuit converter fault diagnosis method to carry out the experiment, and collect the current signal data of five different working states simulated in the experiment, 30 groups of data are collected under each working state, 24 groups of 30 groups of data are selected as the training set, and the remaining 6 groups are selected as the test set. Under the experimental design of the above formula, the experimental results of the two methods are statistically analyzed.

### 3.2 Experimental Results and Analysis

Under the above experimental conditions, the current feature classification results of the traditional electronic circuit converter fault diagnosis method and the diagnosis method designed in this paper are obtained respectively (Fig. 2):

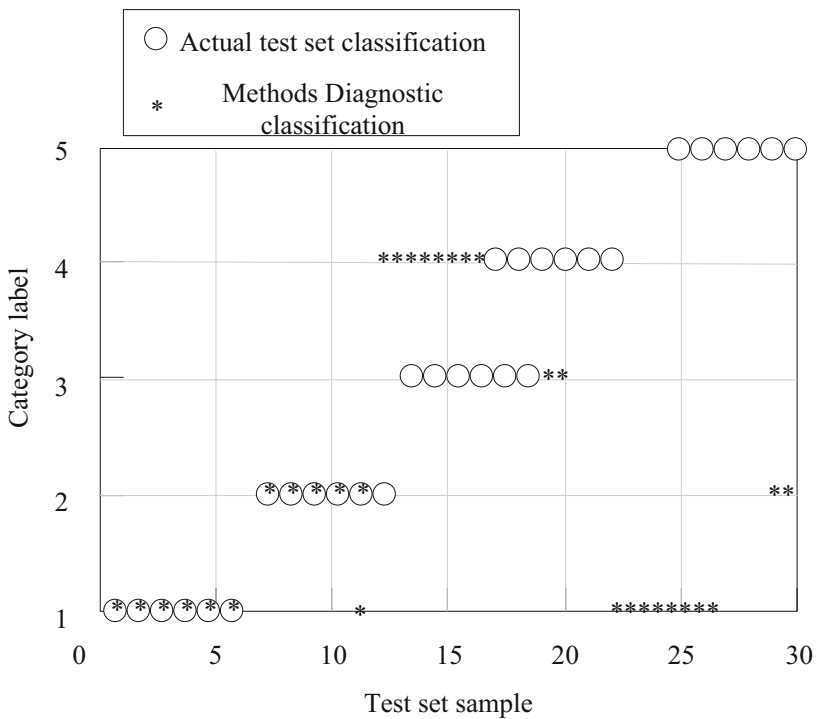


Fig. 2. Feature classification results of traditional methods

The test results of the electronic circuit converter fault detection method based on dynamic sequence response designed in this paper are as follows (Fig. 3):

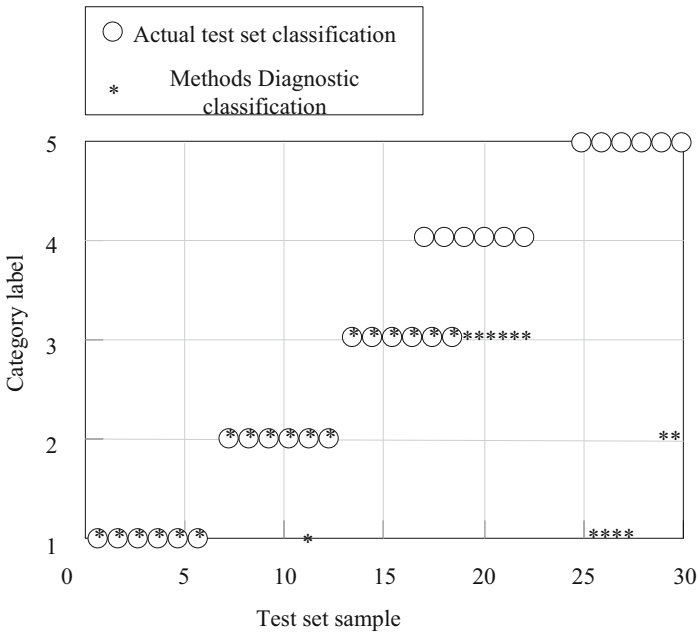


Fig. 3. Feature classification results of the method in this paper

It can be seen from the above two figures that the accuracy of fault identification and classification in the traditional method is only 25.3%, while the accuracy of fault identification and classification in this method can reach 88.7%, so from the accuracy of classification From the above, using the method in this paper can more fully diagnose the overall characteristics of the electronic circuit in the working process, and it is more reliable to classify and evaluate the types of current faults.

### 4 Concluding Remarks

In order to improve and enhance the efficient intelligent diagnosis method of power electronic converter circuit open circuit fault, aiming at the shortcomings of existing fault diagnosis methods in nonlinear and multi class classification problems, this paper deeply studies the open circuit fault diagnosis method of power electronic converter circuit, and proposes a detection method of electronic circuit converter fault based on dynamic sequence response. Firstly, the data collector is designed to complete the collection of electronic circuit converter signal, the analysis model based on dynamic sequence response analysis method is established, the short-circuit fault signal feature extraction based on wavelet transform, the fault detection classifier is optimized, the

fault location algorithm is optimized, and the design of electronic circuit converter fault detection method based on dynamic sequence response is completed. The experimental results show that the accuracy of the traditional method is only 25.3%, and the accuracy of the designed method can reach 88.7%, which verifies the effectiveness of the design method.

At present, power electronic converter circuit fault diagnosis technology is still in the theoretical research stage, there are many aspects need to be further improved. In this paper, some work has been done on the fault simulation and diagnosis method of power electronic converter circuit, but there are still some work that is not deep enough and needs further research and exploration.

**Fund Projects.** 2020 Guangxi University Young and middle-aged teachers' scientific research basic ability promotion project: Research on classroom teaching dynamic evaluation system based on mobile terminal (2020ky45011).

## References

1. Zhao, X., Liu, D.: Analog circuit fault diagnosis and comparison analysis based on SVM optimized by improved fruit fly optimization algorithm. *J. Electron. Meas. Instrum.* **33**(03), 78–84 (2019)
2. Wang, X., Gao, X., Shi, L.: Fault diagnosis analysis of power electronic circuits based on WOA-PNN algorithm. *Microelectronics* **50**(02), 232–235 (2020)
3. Xiong, K., Yue, C., Liu, D., et al.: Analogue circuit fault diagnosis based on SVM optimized by IGWO. *Microelectron. Comput.* **036**(001), 16–21 (2019)
4. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* **458**, 615–625 (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>
5. Sun, J., Hu, G., Deng, W., et al.: Analog circuit soft fault diagnosis based on RS-PSO-SVM integration classifier. *Microelectronics* **50**(02), 227–231 (2020)
6. Gong, Y., Ji, B., Li, J.: Design of floating plate modulator and fault detection circuit for radar transmitter. *Chin. J. Electron Devices* **1**, 20–24 (2020)
7. Yang, S., Xiang, W., Wen, J.: A fault protection scheme based on the difference of current-limiting reactor voltage for overhead MMC based DC grids. *Proc. CSEE* **40**(04), 1196–1211 +1411 (2020)
8. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
9. Guo, X.-P., Liu, S.-Y., Li, Y.: Fault detection of multi-mode processes employing sparse residual distance. *Acta Automatica Sinica* **45**(03), 617–625 (2019)
10. Kong, X., Luo, J., Zhang, Q., et al.: Quality-related fault detection method based on orthogonal signal correction and efficient PLS. *Control Decis.* **35**(05), 1167–1174 (2020)
11. Liu, S., Li, Z., Zhang, Y., Cheng, X.: Introduction of key problems in long-distance learning and training. *Mob. Netw. Appl.* **24**(1), 1–4 (2018). <https://doi.org/10.1007/s11036-018-1136-6>
12. Guo, J., Zhong, L., Li, Y.: Fault detection of multi-mode batch process based on statistics difference LPP. *Appl. Res. Comput.* **36**(01), 123–126 (2019)

13. Yu, J., Zhang, C., Wu, J.: Analog circuit Iddt fault diagnosis combining information entropy in multi-FRFT domain. *Mod. Electron. Tech.* **43**(18), 92–96 (2020)
14. Han, R., Wang, R., Zeng, G.: Fault diagnosis method of power electronic converter based on broad learning. *Complexity* **20**(11), 1–9 (2020)
15. Pandaram, K., Rathnapriya, S., Manikandan, V.: Fault diagnosis of linear analog electronic circuit based on natural response specification using K-NN algorithm. *J. Electron. Test.* **15**(3), 1–14 (2021)
16. Yang, H., Chao, K., Sun, X., et al.: Predictive current control method of photovoltaic energy storage for bidirectional DC-DC converter based on switching sequence. *Power Syst. Technol.* **25**(6), 45–49 (2019)
17. Chu, R., Schweitzer, P., Zhang, R.: Series AC arc fault detection method based on high-frequency coupling sensor and convolution neural network. *Sensors* **20**(17), 49–53 (2020)
18. Xu, S., Tao, S., Zheng, W., et al.: Multiple open-circuit fault diagnosis for back-to-back converter of PMSG wind generation system based on instantaneous amplitude estimation. *IEEE Trans. Instrum. Meas.* **70**(31), 1–13 (2021)
19. Zhao, N., Liu, J., Shi, Y., et al.: Mode analysis and fault-tolerant method of open-circuit fault for dual active bridge DC/DC converter. *IEEE Trans. Industr. Electron.* **87**(9), 131–139 (2019)
20. Shi, J., Deng, Y., Wang, Z.: Analog circuit fault diagnosis based on density peaks clustering and dynamic weight probabilistic neural network. *Neurocomputing* **40**(7), 113–121 (2020)
21. Sheng, Y., Cong, W., Xianghai, B.U., et al.: Detection method of high impedance grounding fault based on differential current of zero-sequence current projection and neutral point current in low-resistance grounding system. *Electric Power Autom. Equip.* **39**(03), 17–22 +29 (2019)
22. Wang, N.: The analysis of electronic circuit fault diagnosis based on neural network data fusion algorithm. *Symmetry* **12**(3), 458–462 (2020)
23. Shi, C., Lu, X.: Online detection method for inter-turn short-circuit fault of permanent magnet synchronous motor based on deep learning. *IOP Conf. Ser. Earth Environ. Sci.* **54**(6), 52–56 (2020)



# Reliability Analysis of Power Side Information Acquisition Model Based on Wireless Sensor

Wen-Lin Xu<sup>1(✉)</sup>, Xin-Ze Guo<sup>2</sup>, Zi-Peng Hu<sup>3</sup>, Chao Li<sup>3</sup>, and Kai Liu<sup>3</sup>

<sup>1</sup> Marketing Department, State Grid Hunan Electric Power Co., Ltd,  
Changsha 410000, China  
dsf2541521@163.com

<sup>2</sup> Changsha Electric Power Technical College, Changsha 410131, China

<sup>3</sup> State Grid Hunan Integrated Energy Service Co., Ltd,  
Changsha 410000, China

**Abstract.** In order to improve the remote monitoring capability of power users' power consumption information, a reliability analysis method of the power side information collection model based on wireless sensor is proposed. Through collecting and screening the operating data of the power side information collection model, detecting the running state of the information on the power side, and modifying the abnormal information by using the adaptive distributed characteristics of the wireless sensor nodes, The vector quantization method is used to process the information fusion of data acquisition, and the attribute classification of power user information is classified by pattern recognition technology. The quantitative fusion tracking method is used to mine and modify the power user information, so as to improve the reliability of the power side information acquisition model. The test results show that the reliability analysis method of the power side information acquisition model based on wireless sensor has higher accuracy in the practical application.

**Keywords:** Wireless sensor network · Power side · Information distribution · Information collection

## 1 Introduction

With the continuous development of power grid, power information intelligent management technology is also improving. The research on the reliability of information collection model based on wireless sensor is of great significance to realize the optimal management and control of electric power and improve the intelligent dispatching performance of electric power information for power users [1]. Research on reliability optimization method of power side information acquisition model based on wireless sensor has attracted much attention. In order to improve the intelligent management and service level of power user information, it is necessary to collect and identify the power user information intelligently, improve the accuracy and reliability of the identification results, and realize the effective monitoring of power user information combined with wireless. Sensor network technology, using adaptive information scheduling method to improve the reliability of power information management and control [2].

This paper studies the collection results of power users' power consumption information, combined with the design technology of wireless sensor network, optimizes the reliability of power users' power consumption information. The remote data acquisition method is adopted to effectively collect the power consumption information of power users, and the collected power consumption information is input into the intelligent information processing terminal to realize the remote reliability control and analysis of power users. Consumption information, improve the real-time processing ability of power users' telecommunication information. It shows the reliability and superiority of the model.

## **2 Reliability of Distributed Data Acquisition Model in Power Side**

### **2.1 Reliability Information Collection of Distributed Information Collection Model in Power Side**

According to the different types of users and business requirements, the comprehensive collection of power reliability information is realized. At present, the wireless sensor network and single-chip microcomputer control method are mainly used to collect the reliability information of power users. The collected reliability information of power users is intelligently stored in the data memory. Combined with big data management technology, the power user information collection and information management are carried out, Improve the reliability of power users' information and intelligent dispatching ability. Aiming at the disadvantages of traditional methods, this paper proposes a dynamic acquisition method of power users' power consumption information based on wireless sensor network reliability information recognition, constructs a wireless sensor network reliability model of power users' power consumption information acquisition, and carries out the adaptive distributed optimal positioning design of power users' power consumption information acquisition sensor nodes [3]. The vector quantization method is used for the reliability fusion processing of the collected data information, and the pattern recognition technology is used for the classification processing of the reliability feature attributes of the power users' power consumption information. The quantitative fusion tracking method is used for the power users' power consumption information mining and feature extraction to realize the automatic mining of the power users' power consumption information, and then the reliability feature design and development of the model are carried out, Finally, the test and analysis of information collection are carried out. The current distributed collection model of reliability information on the power side still has the problems of low utilization rate of a large number of power consumption information data in the practical application process, and the information of power supply enterprises at the provincial, municipal and county levels can not be fully interconnected and shared, which is easy to become a data island phenomenon, It can not meet the current design goals of data fusion and model reliability [4]. In order to solve the above problems, it is necessary to optimize the reliability category of power consumption information collection data (Table 1).

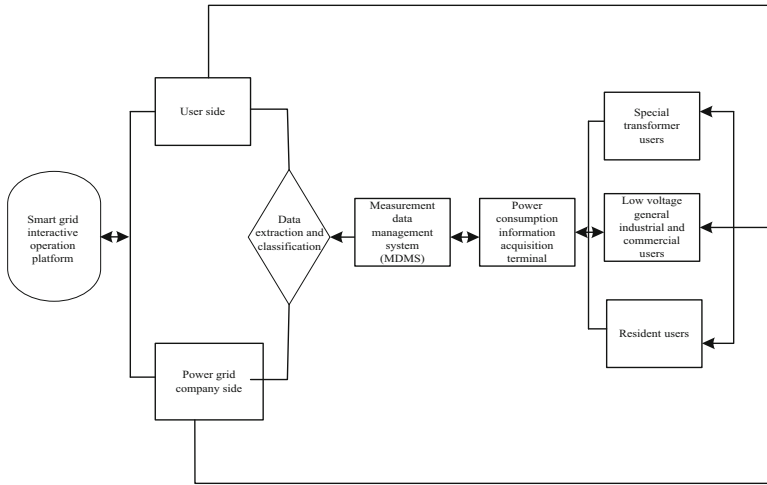


**Table 1.** Reliability index of power consumption information collection data

System name	Total number of systems	Coefficient table of provincial companies	Proportion of provincial companies	City company coefficient table	Proportion of city companies	Number of district and County company systems	Proportion of district and county companies
Centralized meter reading system	248	2	0.40%	210	55.02%	112	44.58%
Load management system	268	7	3.00%	138	78.65%	50	18.35%
Power consumption information acquisition system	1	1	...	0	...	0	...
total	518	10	1.93%	348	67.12	162	30.95%

Based on the data in the above table, the power consumption information collection data is coordinated by the master station, terminal and watt hour meter. According to the different main body of the control logic that executes the calculation of the user's remaining electricity charge and sends out the control tripping command, the way to realize the charge control includes three modes: the master station implements the charge control, the terminal implements the charge control, and the watt hour meter implements the charge control for reliability processing [5]. With the increasing demand of residents for electricity in China, the construction of distribution network in some areas has been unable to adapt to the economy. At the same time, there are many outlets in the distribution network, and the power consumption environment is relatively complex. It is necessary to adopt modern monitoring methods to collect and intelligently monitor the power consumption data, so as to improve the reliability of the model. According to the results of operation and maintenance intelligent monitoring information collection of power consumption information collection system, it is necessary to integrate the reliability characteristic information of marketing business based on the massive data of power consumption information collection, and use the integrated application of marketing and distribution to conduct real-time monitoring on the reliability status, power quality reliability and equipment operation and maintenance engineering reliability of distribution network equipment operation and maintenance. The power consumption information collection system is the main data source of the monitoring system, and it is also a comprehensive just in time information collection and reliability analysis and processing system integrating modern electronic communication technology and power marketing technology. However, the traditional operation and maintenance monitoring has some problems, such as inaccurate monitoring data, poor system redundancy and so on, which can not achieve efficient monitoring of power consumption information collection. Based on the above problems, the optimization design of reliability analysis method for power consumption

information collection, operation and maintenance monitoring based on big data is proposed. This paper analyzes the reliability of power consumption information collection, information preprocessing, information management and information analysis, analyzes the reliability characteristics of database and telecommunication, optimizes the reliability model and active early warning function of power consumption information service according to the characteristics of users, and maintains good reliability performance. According to different cost control methods, the corresponding terminal equipment and electric energy meter are selected for data processing. The electric energy meter includes remote charge control electric energy meter, local charge control electric energy meter and electric energy meter without charge control function, so as to realize data acquisition and improve the reliability of data acquisition. Based on this, this paper optimizes the information reliability acquisition process of intelligent power side, as shown in Fig. 1



**Fig. 1.** Power side reliability information acquisition process

Based on the above steps, it is judged whether the user ends the electricity management behavior, and the time is based on the instantaneous stop value of the electric equipment. However, in the actual situation, the normal operation of electrical equipment through human or its own characteristics, there is a pause time gap, which may cause error to judge the end time [6]. Based on this, the maximum interval time parameter is considered to solve the problem of poor reliability of the model. When the power consumption behavior ends, it will wait for a period of time for the next calculation and reasoning, and the waiting time for data collection will be less than TIG. If the power consumption behavior does not conduct data screening again in the TIG period, it will automatically determine the end time of equipment operation [7]. Because different electrical equipment have their own operating characteristics, it is necessary to extract the corresponding operation scenario information from the basic

electricity consumption information data automatically. The characteristics of electrical equipment are extracted from the ontology of electrical equipment, and the tasks are classified by the activity characteristics of electrical equipment. The reliability characteristics of the model are also obtained from the ontology, so as to ensure the research effect.

### 2.2 Reliability Algorithm of Information Distribution in Power Consumption Side

In order to achieve the reliability of power user information collection, the power user information distributed wireless sensor network structure model is used to design the array network structure. The power user information collection wireless sensor network array adopts the distributed Internet of things networking design method, and the sensor nodes are arranged in the power user information monitoring area, The multi-sensor quantitative detection and recognition method is used to collect the dynamic characteristics of power user information

$$x(k + 1) = A(k)x(k) + \Gamma(k)w(k) \tag{1}$$

$$z_i(k) = H_i(k)x(k) + u_i(k)x(k + 1) \tag{2}$$

Where:  $x(k) \in R, n \times 1$  is the reliability matrix of monitoring state information of power users,  $A(k) \in R, n \times n$  is the reliability characteristic distribution matrix of non-stationary power users' power consumption information, the background interference noise of information collection is  $w(k)$ ,  $u_i(k)$  is a Gaussian white noise with zero mean value and variance  $u$ , and  $\Gamma(k)$  is the reliability sample correlation matrix of dynamic sensing information of power users' power consumption information [8].  $z_i(k) \in R, p \times 1$  is the mixed data of power user reliability information collected by the  $i$ -th wireless sensor network node,  $H_i(k) \in R, p \times n$  is the impulse response of wireless sensor network information collected by power user information, where the interference noise  $u_i(k) \in R, p \times 1$  of data collection process is Gaussian white noise with zero mean value and variance  $D_i(k)$ . Based on the assumption that there are association rules between the weighted matrix  $w(k)$  and the reliability matching vector  $u_i(k)$ , the fuzzy information weighted control technology is used to process the multi-mode feature reliability of power consumption information, and the multi-sensor fusion tracking identification method is used to obtain the node distribution structure expression of reliability information collection

$$\begin{cases} E\{W(k)z_i(k)/u_i^T(k)\} = B_i(k), i = 1, 2, \dots, N \\ E\{u_i(k)u_j^T(k)\} = D_{ij}(k), i, j = 1, 2, \dots, N, i \neq j \end{cases} \tag{3}$$

Considering the sensor network composed of  $N$  sensor nodes, the output structure model of power user electricity information sampling network is constructed by using sensor node array, and the node location of power user electricity information data acquisition sensor network is obtained

$$\begin{cases} x_{i,t+1} = (x_{i,t} + x'_{i,t+1})/2B_i(k) \\ y_{i,t+1} = (y_{i,t} + y'_{i,t+1})/2D_{ij}(k) \end{cases} \tag{4}$$

The feedback equalization filter is constructed based on the average value of the measurement information of wireless sensor network nodes. In  $t + 1$  monitoring node, the coordinates of the origin moment  $o_i$  of the monitoring data distribution of the power user reliability information are obtained as  $(x_i, t + 1, y_i, t_i + 1)$ . the multi-sensor network is used to integrate the reliability information of the remote monitoring of the power user power consumption information, The dynamic statistical characteristics of the monitoring data of power user reliability information collected by output are as follows:

$$E[\bar{V}(k)] = \Delta E \tag{5}$$

$$E[x_{i,t+1}\bar{V}(k) - y_{i,t+1}\bar{V}^T(k)] = R(k) \tag{6}$$

The quantitative noise  $w(k)$  and  $\bar{V}(k)$  of power user information collection are not related to each other. The information fusion processing of power user information is carried out by combining reliability information matching detection method. Frequent item detection and association rule feature extraction are carried out for high-dimensional fusion data, thus the association rule feature extraction and information fusion processing of power user power user information are realized. In order to realize the optimal collection of power user information, it is necessary to build a distributed data structure model of power user power information under wireless sensor network mode. A quad  $G$  is used to represent the fuzzy distributed structure storage center of power user power information. It is  $C$ . suppose  $I$  is the phase space embedding dimension of the interaction of power user information, The information flow model of the power user information expressed by the constraint equation of the association rule term is as follows:

$$R(k)\Delta E + E[\bar{V}(k)\bar{V}^T(k)] = \Delta R(k) \tag{7}$$

The model reliability evaluation algorithm also plays an important role in the design of power consumption information collection. In order to ensure the reliability of the model, it is necessary to collect the power consumption information and calculate the reliability under normal conditions

$$W(a)_{\min} = \Delta R(k) \sum_{i=1}^n \lambda_i (k_i^t a) + \delta y(a) \tag{8}$$

The reliability judgment function of information collection can be expressed by formula:

$$W(\mathbf{b})_{\min} = \sum_{i=1}^m \lambda_i(-b_i) - \delta y \left( \frac{1}{\delta} \sum_{i=1}^m k_i b_i \right) \tag{9}$$

Where:  $k_1, \dots, k_m \in 2^m$  is the vector of  $m$  information;  $\lambda_1, \dots, \lambda_m$  is the loss quantity of reliability information;  $y$  is the reliability coefficient of information collection;  $\delta$  is a conjugate function;  $a$  is the original information variable;  $b$  is the variable of distributed information function;  $t \geq 0$  is the function parameter. Each information variable  $c_i$  corresponds to the  $D_i$  of a storage variable

$$\lambda_i(c_i^t a) = \frac{(c_i^t a - D_i)^2 \cdot y(a)}{2} = |a|_i \tag{10}$$

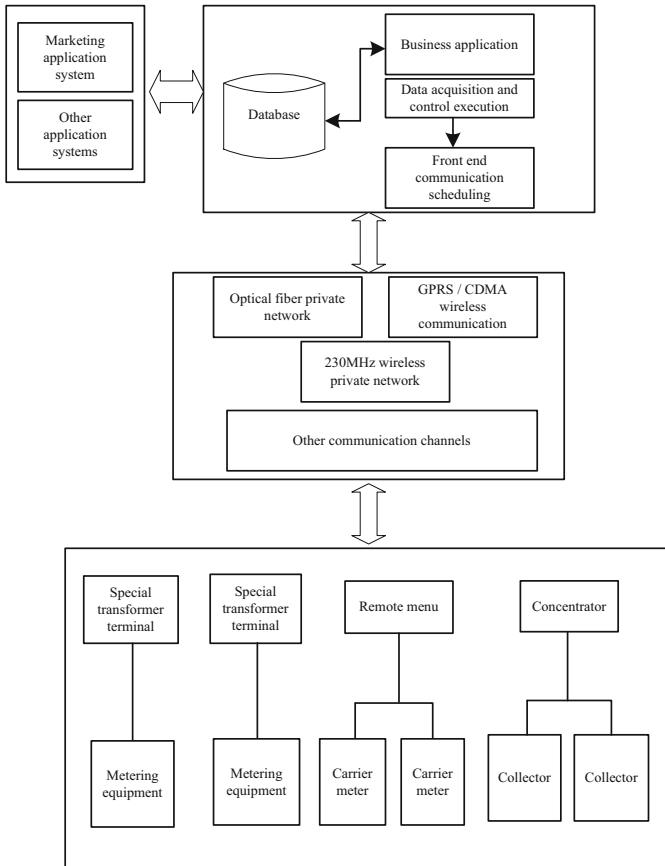
The constrained function model can be obtained from the formula

$$f(a)_{\min} = \|Fa - r\|_3^3 + W(\mathbf{b})_{\min} t |a|_i \tag{11}$$

Where:  $F$  belongs to the matrix of  $n$  term distributed information;  $t > 0$  is the parameter of information constraint. Based on the above algorithm for power side information distribution data acquisition and processing, can better improve the effect of data acquisition.

### 2.3 Reliability Optimization of Distributed Data Acquisition Model in Power Side

The purpose of reliability design of distributed information collection model based on overall architecture is to make the collected information accurate. In the environment of Internet of things, information collection has the characteristics of strong distribution and wide range. Therefore, the design needs to be divided into different levels: storage layer and access layer [10]. The storage layer can store a certain amount of distributed information of power equipment. After information collection, the command is given to the collector, and the collected information is combined with logical thinking to accurately determine the distributed location of power equipment. Through the determination of location information, the command is sent to the access layer, so as to realize the distributed information collection in the Internet of things environment, Acquisition logic architecture is mainly from the logical point of view of electricity information acquisition management from the master station, channel, terminal, acquisition point and other four levels of reasonable logic classification, in order to ensure the reliability of the model. The logic architecture of power consumption information collection and management is shown in Fig. 2



**Fig. 2.** Structure framework of distributed information collection model at power consumption side

The management of power information collection can be divided into three levels logically: the main station layer, the communication channel layer and the acquisition equipment layer. Besides the reliability of marketing information, it is called other application information. Through the interface, acquisition management can effectively connect and screen with marketing applications and other applications. The main station layer is composed of business application, data acquisition, control execution, pre communication scheduling and database management. Data acquisition is responsible for collecting user power information and protocol analysis, so as to improve the reliability of model motion. The application of distributed data collection model of electric side information is used to realize the optimization goal of logic reliability of various applications. The communication management and scheduling of various terminals are carried out through remote communication, and the related control operations are performed on terminals with control function. Communication network is the carrier of information interaction. The main ways of connecting the main

station and collecting and transmitting table network are cable communication, optical fiber communication and carrier remote communication information management structure as shown in Fig. 3

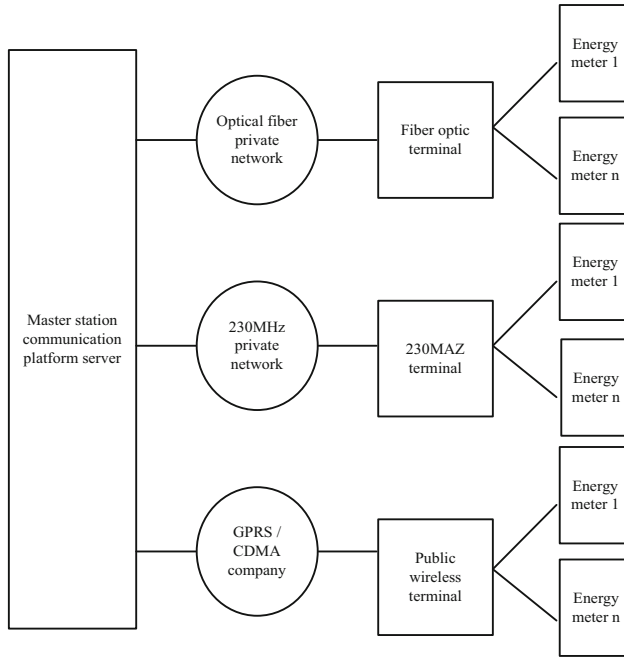
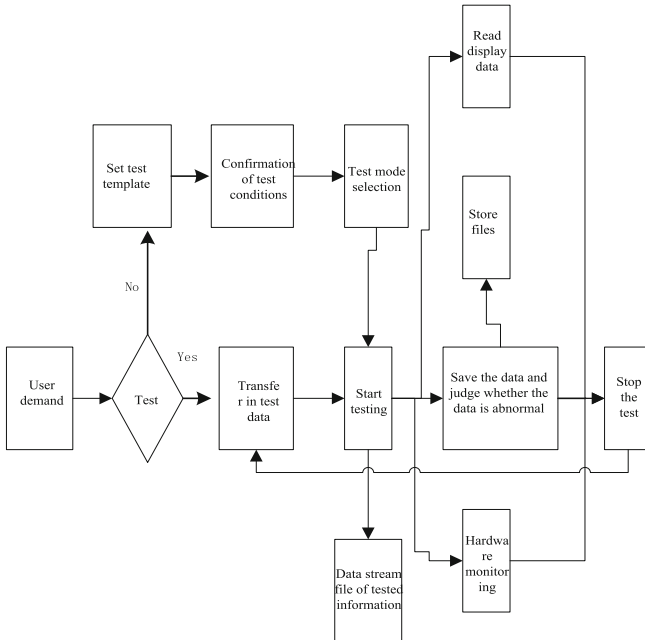


Fig. 3. Description of remote and local information processing

Because the data needed for power consumption information collection, operation and maintenance monitoring is real-time, the system needs to quickly and accurately judge the operation position, intelligently diagnose all data, model information, graphical information and a series of unstructured data, and process the data reliably in real time. The model structure is used to analyze whether the change of charge characteristics in distribution transformer area meets the monitoring standard. Combined with the load and circuit monitoring in distribution transformer area, the reliability of power distribution under the line is analyzed. HBase is used to store the data of power consumption information and analyze it quickly and reasonably to improve the operation reliability of the model. Based on part of the difference information, it needs to be optimized in combination with low-voltage power line carrier communication technology. The most important way of local communication for power consumption information collection is relatively mature technology development. The industry has a certain scale and can support large-scale construction of power consumption information collection. Low voltage power line carrier communication has incomparable advantages over other communication technologies, One is to use low-voltage power line as communication medium to naturally form information distribution network with

clear boundary, which is conducive to the overall planning of network layout of remote channel and local channel. In the design of data acquisition, the user information and equipment information are managed macroscopically according to the user's requirements; At the micro level, the sampling time is designed. The design is carried out in strict accordance with the real business process of data acquisition under the state of security environment. The order is: equipment reliability correction, reliability parameter setting, reliability data storage, reliability data calculation, data display and export. The specific steps are shown in Fig. 4.

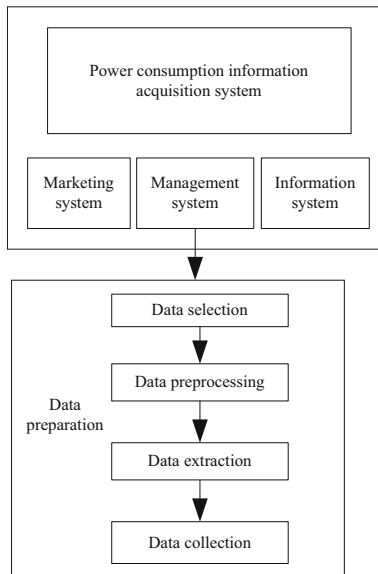


**Fig. 4.** Power consumption information reliability management process

The collection and monitoring of reliability characteristics of power consumption information is established according to the characteristics of users and the situation of distribution network users. According to the established business analysis model library, different kinds of models are designed, including market development analysis, multi-dimensional analysis, line damage analysis, power quality analysis, operation and maintenance pre control analysis, work efficiency analysis and user recovery risk analysis. Different business models have different data sources and algorithm libraries, and each information integration and publishing needs to be combined with the actual application scenarios to complete the business model design. In the big data environment, the design of power consumption information collection, operation and maintenance monitoring system is based on the assumption of market development and business expansion.



1. ① Obtain the capacity, voltage level, power supply address, charge characteristics and other information of the capacity increasing transformer;
2. ② By locating the position of power supply station, the operation and distribution integration of lines, switching stations and stations is carried out;
3. ③ According to the data obtained from electricity statistics and the charge of each dimension of the line, combined with the marketing business process and power grid model information, the big data analysis method is introduced to monitor the new power supply location. The model design is shown in Fig. 5.



**Fig. 5.** Power consumption information business information management model

In this model, there are relational database and non relational database, relational database is the mainstream database, which is widely used in system design. The advantage of this method is that it can collect and store the power consumption information by using two-dimensional structure storage mode, and its rich integrity reduces the probability of data inconsistency, which is suitable for dealing with relatively complex online analysis system. In the current big data environment, due to the influence of I/O hard disk, the problems of difficult expansion and complex query, the use of relational database can solve the problem of unstructured data storage [5]. Non relational database has the advantages of high reliability and strong scalability, which is suitable for distributed storage system. It follows the principle of data storage, stores data with key value, and the structure is uncertain, which makes the stored data not limited to fixed structure, and saves monitoring time. The main functions of the model are as follows: to initialize the data acquisition and set up each acquisition channel;

Setting the collection environment and parameter information; Control the automatic process of data acquisition; In the process of acquisition, the data of each acquisition channel needs to be stored, calculated, analyzed and displayed to be processed synchronously; Strict management of user groups, improve security and confidentiality. In the design process, it is set to record the change of centralized information communication operation data every one minute, the time interval is 50 ms, and the upper limit of acquisition cycle is 10 min. when each collector is connected, it is easy to lose 10 data [8]. Therefore, we need to use multi port collector to collect the lost data.

Based on the wireless sensor network model of power user information collection and the adaptive distributed optimal positioning design of power user information collection sensor nodes, the optimal design of power user information dynamic collection technology is carried out. The distribution function description of power user information distribution structure model is as follows

$$X_p(u) = \begin{cases} p\sqrt{\frac{1-j\cot\alpha}{2\pi}}e^{\frac{1}{2}\cot\alpha} \int_{-\infty}^{+\infty} x(t)e^{\frac{1}{\sec\alpha t - j\sin\alpha}dt}, \alpha \neq n\pi \\ x(u), \alpha = 2n\pi \\ x(-u), \alpha = (2n \pm 1)\pi \end{cases} \tag{12}$$

Where: p is the characteristic order of the reliability structure of the power information storage of distributed power users, and the reliability regression analysis model of big data to be mined is constructed:

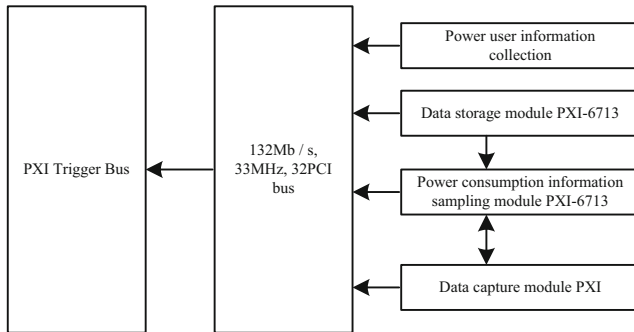
$$\begin{cases} \min \sum_{1 \leq i \leq Ke} \sum_{Ck|e} \frac{f(e(i))}{C(e,i)} \\ 0 \leq f(e, i) \leq C(e, i) \\ F = \text{const} \\ \sum_{1 \leq i \in K, e \subseteq k(e)} \frac{f(e(i))}{C(e,i)} + \sum_{e \subseteq k|e} \frac{f(e'(i))}{C(e',i)} \leq k(v) \end{cases} \tag{13}$$

A multivariate statistical characteristic equation is used to describe the reliability information of power users

$$\begin{pmatrix} X \\ P(X) \end{pmatrix} = \begin{cases} a_1, a_2, \dots, a_m \\ p(a_1), p(a_2), \dots, p(a_m) \end{cases} \tag{14}$$

In order to realize the accurate and real-time acquisition of power users' power consumption information, the embedded technology is used for reliability information acquisition and development design. VXI bus scheduling and embedded information reliability acquisition technology are used for wireless sensor network structure design and reliability adaptive information processing of power users' power consumption information, PXI trigger bus is used to control the clock sampling reliability of power users' power consumption information to improve the integrity and reliability of power users' power consumption information collection. Data playback is carried out in the application layer module to realize the dynamic monitoring of power users' power

consumption information, The overall structure of the power user reliability information collection designed in this paper is shown in Fig. 6.



**Fig. 6.** Overall structure optimization of power user information acquisition system

Power users' power consumption information collection uses multi-sensor distributed in the power users' power consumption information monitoring area for data collection, and uses wireless sensor network networking technology for sensor networking reliability design. Wireless sensor network networking is divided into physical layer, data link layer and data transmission layer. In order to ensure the security and reliability of the information collection model, the application service layer of wireless sensor network for power user information collection is constructed, the reliable scheduling of data bus transmission is carried out, and the multi-mode embedded control method is used for adaptive transmission and reliability control of power user information.

### 3 Analysis of Experimental Results

In order to verify the reliability of the distributed collection model of power consumption information based on the wireless sensor network, experiments were carried out. First, set the experimental parameters based on the overall structure of the distributed information collection design framework to provide data for the experiment; select a power company, and let the company employees upload the distributed information of the power equipment at the same time, record the information collection speed in time, and conduct data analysis; The quantized noise  $w(k)$  and  $V(k)$  collected by power user electricity information are not correlated with each other, and the matching correlation detection method is used for information fusion processing of power user electricity information, and frequent item detection and correlation of high-dimensional fusion data Rule reliability feature extraction, which realizes the association rule feature extraction and information fusion processing of power users'

electricity consumption information. The experiment uses hierarchical iterative collection of multi-source feature digital information as the test environment of the experimental group, and Deep Web collection as the test environment of the control group. With the support of the Google Play platform, the change trend of the coverage area of reliability information and the iterative drive loading capacity is recorded. As shown in Table 2.

**Table 2.** Experimental data

Acquisition frequency (Hz)	Duration (s)	Number of experiments (Times)	Overflow or not
500	2400	3	no
800	1130	3	no
1200	850	3	no
1500	150	3	yes
110000	15	3	yes

According to the experimental conditions and the experimental environment in the table, the data is read every one minute. Based on this, the experimental equipment and parameters are further standardized and designed, as shown in Table 3.

**Table 3.** Experimental equipment and parameters

Parameter	Upper computer	Detector
Number of equipment	1 set	10 sets
Sampling interval	1s	80ms
Cycle/ms	5000	7000
Actual time	500000 ms	(data loss)

Based on the above experimental environment, the main test end organization, experimental group end organization, and control end organization are open at the same time, making the Google Play platform enter a stable iterative operation state. Under the condition that the number of icons in the display interface box remains unchanged, open OriginPro2018 and use the built-in numerical statistics function to record the experimental indicators. According to the specific experimental results, the experimental group and the control group are analyzed and applied, and then the digital information is iteratively driven on the basis of the load. According to the change of capacity, the distribution of electricity consumption information is collected on this basis, as shown in Fig. 7. Show.

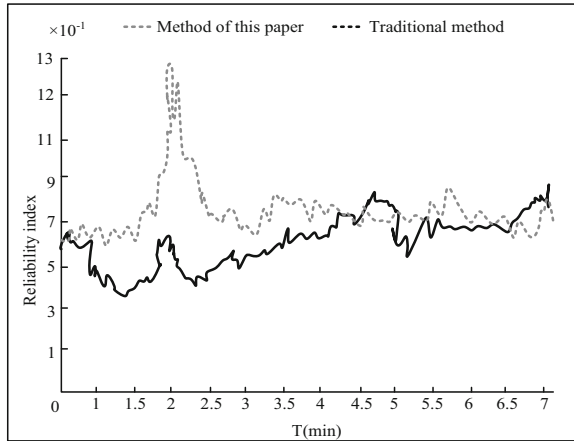


Fig. 7. Comparison diagram of model reliability index changes

In order to test the design performance of the power user electricity information collection designed in this paper, a simulation experiment was carried out. In the experiment, 32 sensor nodes were used to design the reliability effect of the power user electricity information collection. The low trace noise of the data collection: 0.00 4dB/rms, frequency option: from 9 kHz/100 kHz (with biased T-type connector) to 4.5 GHz/8.5 GHz, configure VisualDSP, perform hardware simulation, set two D/A channels to output a maximum of 5V at the same time, power users use electricity The pulse width of the information output is 2 μs. According to the above parameter setting, the power user electricity information collection is performed, and the stability and collection accuracy of the test are tested. The comparison result is shown in Fig. 8.

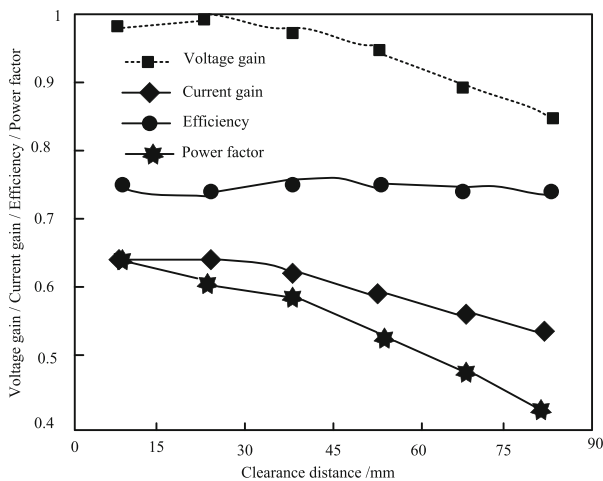


Fig. 8. Reliability detection of electricity consumption information distribution collection model

Analyzing Fig. 8 we know that the automatic control performance designed in this paper for the collection of power user electricity information is better, and the collection results are accurate and reliable.

## 4 Concluding Remarks

Combining wireless sensor network technology and realizing the effective monitoring of users' electricity consumption information according to the results of user electricity consumption information collection, the adaptive information dispatching method is adopted to improve the power information management level and the reliability of the model. A method for dynamic collection of electricity consumption information of electric power users based on distributed array wireless sensor network network and quantitative detection and identification is proposed, and a wireless sensor network model for collecting electricity consumption information of electric power users is constructed. The acquisition sensor is used for the adaptive distributed optimization positioning design of the node, the vector quantization method is used for information fusion processing on the collected data, the pattern recognition technology is used for the attribute classification processing of the electric energy information of the power user, and the quantitative fusion tracking method is used for the power user's Electric energy information mining and feature extraction have realized the reliability of automatic mining of user electricity information. Studies have shown that the reliability and reliability of collecting electricity consumption information from electricity users is high.

## References

1. Gombé, B.O., Mérou, G.G., Breschi, K., et al.: A SAW wireless sensor network platform for industrial predictive maintenance. *J. Intell. Manuf.* **30**(4), 1617–1628 (2019)
2. Gupta, O., Goyal, N., Anand, D., et al.: Underwater networked wireless sensor data collection for computational intelligence techniques: issues, challenges, and approaches. *IEEE Access*, **PP**(99), 1–1 (2020)
3. Lorenz, P., Schott, R., et al.: New path centrality based on operator calculus approach for wireless sensor network deployment. *IEEE Trans. Emerging Topics Comput.* **7**(1), 162–173 (2019)
4. Singh, A., Nagar, J., Sharma, S., et al.: A Gaussian process regression approach to predict the k-barrier coverage probability for intrusion detection in wireless sensor networks. *Expert Syst. Appl.* **172**, 114603 (2021)
5. Zhou, G.D., Xie, M.X., Yi, T.H., et al.: Optimal wireless sensor network configuration for structural monitoring using automatic-learning firefly algorithm. *Adv. Struct. Eng.* **22**(4), 907–918 (2019)
6. Chen, X., Zhang, X.: Secure electricity trading and incentive contract model for electric vehicle based on energy blockchain. *IEEE Access* **7**(10), 178763–178778 (2019)
7. Yu, K., Guo, N., Cao, Z., et al.: Fast information acquisition using spectra subtraction for Brillouin distributed fiber sensors. *Optics Express* **27**(7), 9696 (2019)

8. Li, H., Zhao, G., Qin, L., et al.: Design and optimization of a hybrid sensor network for traffic information acquisition. *IEEE Sens. J.* **20**(4), 2132–2144 (2020)
9. Lu, W., Liu, G., Si, P., et al.: Joint resource optimization in Simultaneous Wireless Information and Power Transfer (SWIPT) Enabled Multi-Relay Internet of Things (IoT) system. *Sensors* **19**(11), 2536 (2019)
10. Liu, C.-H., Hsu, C.-S., et al.: Fundamentals of simultaneous wireless information and power transmission in heterogeneous networks: a cell-load perspective. *IEEE J. Sel. Areas Commun.* **37**(1), 100–115 (2019)



# Research on Obstacle Avoidance Tracking Planning of Hyper-redundant Manipulator Based on VR Technology

Xiao-zheng Wan<sup>1(✉)</sup>, Song Zhang<sup>2</sup>, Ji-ming Zhang<sup>1</sup>, Hui Chai<sup>1</sup>,  
and Dong-bao Ma<sup>3</sup>

<sup>1</sup> Institute of Oceanographic Instrumentation, Qilu University of Technology  
(Shandong Academy of Sciences), Qingdao 266061, China  
wanxzz302@outlook.com

<sup>2</sup> College of New Energy, China University of Petroleum,  
Qingdao 266580, China

<sup>3</sup> College of Mechatronic Engineering,  
Beijing Polytechnic, Beijing 100176, China

**Abstract.** In order to ensure that the super-redundant manipulator can work without collision in the obstacle space, this paper designs an obstacle avoidance trajectory planning method for the super-redundant manipulator based on VR technology. The super-redundant manipulator with seven degrees of freedom was selected as the research object, VR technology was introduced to construct the virtual model of the super-redundant manipulator, and the redundancy space of the manipulator was determined according to the model. Then, the pre-selected minimum distance index is selected as the obstacle avoidance index of the mechanical arm. Based on the calculated real-time minimum distance, the inverse kinematics equation of the mechanical arm is obtained by using the escape velocity dynamic obstacle avoidance algorithm, so as to realize the tracking planning of the obstacle avoidance trajectory of the super-redundant mechanical arm. Simulation experiment results show that: it can be seen from the curve of the joint Angle and angular velocity of the manipulator that the proposed method successfully avoided obstacles and reached the target configuration after the application of the proposed method, which met the requirements of the obstacle-avoidance trajectory tracking planning of the current hyperredundant manipulator, and fully proved the effectiveness of the proposed method.

**Keywords:** VR technology · Super redundant manipulator · Obstacle avoidance trajectory planning · Inverse kinematic equation of mechanical arm

## 1 Introduction

As a kind of joint mechanical device simulating the human arm, the mechanical arm has developed into the supporting technology of modern high-end manufacturing industry, and has been widely used in many fields such as material welding, spraying, handling, cutting and so on. Through the operation of the manipulator arm, not only the



production efficiency and operation quality is improved, but also the existing labor environment is improved, so that it can gradually replace the manual operation in high-risk environment [1]. With the continuous expansion of production requirements, the operating environment of manipulator in many application fields is no longer a relatively free space, but a constrained space formed by many obstacles. Therefore, the obstacle avoidance planning of manipulator has become one of the important research directions in the technical field of manipulator [2].

For hyperredundant manipulator, the number of degrees of freedom of the joints is greater than the degree of freedom limited by the end-effector pose. Each joint can move from an initial joint position to the desired joint position in joint space without changing the end-effector position (pose). Because of this feature, the hyperredundant manipulator improves the joint configuration of the low-DOF manipulator and overcomes the shortcomings of the general manipulator, such as poor flexibility, low obstacle avoidance ability, joint overrun and poor dynamic performance, which has become the focus of people's attention.

The redundancy of the hyperredundant manipulator brings difficulties to its inverse motion process. One pose of the hyperredundant manipulator corresponds to numerous joint configurations, and how to choose from the numerous configurations becomes a difficulty. When hyperredundant manipulator moves in free space, trajectory tracking planning can be accomplished by using inverse kinematics. However, when the hyperredundant manipulator moves in the obstacle space, it is very important to plan a collision free trajectory for the safety of the hyperredundant manipulator and other equipment in the space.

At present, relevant scholars have designed a series of obstacle avoidance trajectory planning methods for manipulator, such as the obstacle avoidance trajectory tracking planning method for hyperredundant manipulator based on particle swarm optimization algorithm and the obstacle avoidance trajectory tracking planning method for hyperredundant manipulator based on multi-objective ant colony algorithm, etc. However, in practical application, it is found that the ideal degree of application effect of the above traditional methods is low. Therefore, this paper proposes an obstacle avoidance trajectory planning method of super redundant manipulator based on VR technology. The design idea of the new method is as follows:

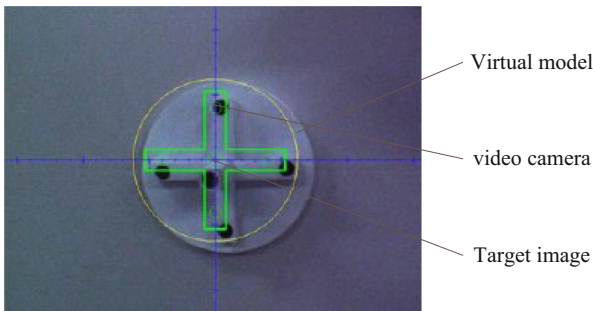
- (1) Taking the 7-DOF hyperredundant manipulator as the research object, the virtual model of the hyperredundant manipulator was constructed by VR technology, and the redundant space of the manipulator was determined according to the constructed model.
- (2) The pre-selected minimum distance index was selected as the obstacle avoidance index of the manipulator, and the real-time minimum distance was calculated. Then, the inverse kinematics equation of the manipulator was established by using the escape velocity dynamic obstacle avoidance algorithm, so as to realize the tracking planning of the obstacle avoidance trajectory of the hyperredundant manipulator.

## 2 Research on Tracking Planning Method of Obstacle Avoidance Trajectory Information of Hyper-redundant Manipulator

### 2.1 Introduction of VR Technology

Due to the human participation, in the past, the accuracy of the obstacle avoidance trajectory planning process of super redundant manipulator is very low, which only uses the naked eye to observe the target. It is high precision and inflexible to solve the problem by machine. According to the data given by the computer, the operator's operation mode combines the advantages of the two well, and VR technology can realize the good combination of the two [3].

In order to facilitate the operator to observe the video scene, the virtual model is drawn in wireframe mode, as shown in Fig. 1.



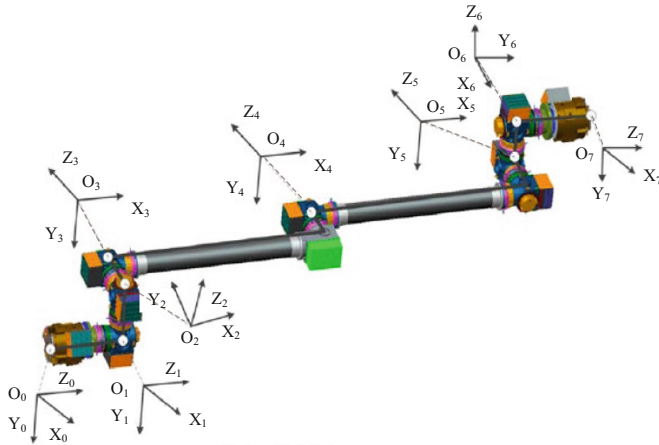
**Fig. 1.** Superposition of target image and virtual model

The video image is captured by the camera carried by the end of the manipulator. The virtual model represents the state that the manipulator moves to the target position accurately, that is, the state that the target and its mark point should move to when it can grab the position of the replaceable module [4].

The camera used in this paper is the industrial camera produced by point grey company, the model is FL2-03S2C, the resolution is  $640 \times 480$ , the lens is 09C Industrial lens of computer company, the field of view is  $32^\circ$ .

### 2.2 Build a Virtual Model of the Robotic Arm

A seven-degree-of-freedom super-redundant manipulator is selected as the research object, which is composed of seven joints, two end effectors, and two arm rods. Its structure is shown in Fig. 2.



**Fig. 2.** Three-dimensional model and coordinate system of a seven-degree-of-freedom super-redundant manipulator

Adopt the degree of freedom configuration of  $R\perp R\parallel R\perp R$ ,  $R$  represents the rotation joint,  $\perp$  represents the axis of the two joints are perpendicular to each other,  $\parallel$  represents the axis of the two joints are parallel to each other. There is an end effector at each end of the robotic arm, and one end effector is used to connect the robotic arm to the space station and serves as the working base of the robotic arm. The other end effector serves as a tool for the robotic arm to capture operations. The base and end functions of the robotic arm are fully reused, so that the robotic arm has a “crawling” function and increases the operating space. Seven degrees of freedom at the same time, the super redundancy configuration, not only can be done using redundant features from the singular, obstacle avoidance, joint torque optimization task, increase the mechanical arm operation flexibility, but also can be used as a backup technology, in the case of a joint failure of mechanical arm, the rest of the six joints still can complete the task of three dimensional space [5].

The virtual model of the robotic arm is established according to the coordinate system of Fig. 1, and the corresponding D-H parameters are shown in Table 1.

**Table 1.** D-H parameter table of virtual model of robotic arm

$i$	$\theta_i$ ( $^\circ$ )	$d_i$ (mm)	$a_{i-1}$ (mm)	$\alpha_{i-1}$ ( $^\circ$ )
1	0	$a_0$	0	0
2	90	$a_1$	0	90
3	0	$a_2$	0	-90
4	0	$a_4$	$a_3$	0
5	0	$a_6$	$a_5$	0
6	-90	$a_7$	0	90
7	0	$a_8$	0	-90

The transformation matrix of 7-DOF super redundant manipulator is as follows:

$${}^0_7T = {}^0_1T_1{}^1_2T_2{}^2_3T_3{}^3_4T_4{}^4_5T_5{}^5_6T_6{}^6_7T_7 \tag{1}$$

Equation (1) represents the forward kinematics model of the robotic arm, and establishes the mapping relationship between the operating space and the joint space. From this formula, the pose transformation from the 7th coordinate system to the base coordinate system  $O_0 - x_0y_0z_0$  can be obtained.

The core problem of hyper-redundant manipulator kinematics is inverse kinematics. Assuming the end pose (task space)  $X = [x_1, x_2, \dots, x_m]$ , joint space  $\theta = [\theta_1, \theta_2, \dots, \theta_n]$ ,  $n > m$  of the hyper-redundant manipulator, there are:

$$X = f(\theta) \tag{2}$$

In formula (2),  $f(\bullet)$  represents the mapping function from joint space to task space, that is, the positive kinematics equation. Knowing the value of each joint variable, the pose of the end can be uniquely determined. When the end pose is known, solving the value of each joint variable is an inverse kinematics problem [6]. Since the dimension of the joint space is greater than the dimension of the task space, the inverse kinematics is solved directly, and the same pose corresponds to an infinite number of configurations. Therefore, it is necessary to supplement appropriate constraints to make the inverse kinematics of the hyper-redundant manipulator have a unique solution. This is the difficulty of the inverse kinematics of the super-redundant manipulator, and it is also the advantage of the super-redundant manipulator. It can improve the working quality of the manipulator or achieve other goals by supplementing appropriate constraints.

Most of the inverse kinematics of hyper-redundant manipulators are solved by velocity-based methods. Knowing the joint velocity  $\dot{\theta}$  and the Jacobian matrix  $J \in \mathbb{R}^{m \times n}$ , the end pose velocity  $\dot{X} = J\dot{\theta}$  can be obtained. Velocity-based inverse kinematics is the inverse process of the above process, that is, given the pose velocity  $\dot{X}$ , the joint velocity  $\dot{\theta}$  is solved, and the joint angle is obtained by integrating the joint velocity. For the hyper-redundant manipulator, under the condition of full rank of the Jacobian matrix,  $\dot{\theta} = J^{-1}\dot{X}$ ; but because the Jacobian matrix of the hyper-redundant manipulator is not a square matrix, its inverse value cannot be calculated.

The scholar Whitney first proposed the pseudo-inverse method to solve the inverse kinematics of the hyper-redundant manipulator. In essence, it is an optimization problem as follows:

$$\begin{cases} \min \|\dot{\theta}\| \\ \text{subject to : } \min \|\dot{X} - J\dot{\theta}\| \end{cases} \tag{3}$$

Equation (3) shows that the norm of joint velocity should be minimized on the premise of minimum error of end tracking trajectory, so the solution obtained by pseudo inverse method is also called minimum norm solution. The Lagrange multiplier method is used to solve the optimization problem of Eq. (3):

$$\dot{\theta} = J^+ \dot{X} \tag{4}$$

In Eq. (4),  $J^+$  represents the pseudo inverse of Jacobian matrix.

### 2.3 Determine the Redundant Space of Manipulator

The redundancy space of the robot arm greatly affects the accuracy of obstacle avoidance trajectory tracking planning. Therefore, the determination of the redundancy space of the robot arm is studied in this paper.

For the hyper-redundant manipulator, the Jacobian matrix  $J$  is a rectangular matrix. If the rank  $r = m$  of the Jacobian matrix, the column space of the Jacobian matrix is the entire  $R^m$  space. For any end velocity  $\dot{X}$ , the Eqs. (5) has an infinite number of feasible solutions  $\dot{\theta}$ , and redundancy refers to the uncertainty of Eq. (5). Assuming that  $\dot{\theta}_s$  is a special solution of Eq. (5), the general solution of Eq. (5) can be expressed as  $\dot{\theta} = \dot{\theta}_s + \dot{\theta}_a$ , then there is:

$$\dot{X} = J(\dot{\theta}_s + \dot{\theta}_a) \tag{5}$$

Obtained by formula (5):

$$J\dot{\theta}_a = 0 \tag{6}$$

Given the terminal velocity  $\dot{X}$ , there are infinitely many feasible solutions of joint velocity  $\dot{\theta}$ , which are composed of the special solution  $\dot{\theta}_s$  and the general solution  $\dot{\theta}_a$  of the corresponding homogeneous equations  $J\dot{\theta}_a = 0$ . This means that  $\dot{\theta}_a$  is the element of the zero space  $N(J)$  of Jacobian  $J$ , and all of these solutions constitute the zero space of Jacobian matrix, which is the  $n - m$  dimensional subspace of  $n$  dimensional real space  $R^n$ , denoted as  $N(J)$ . It is orthogonal to row space  $R(J^T)$  of Jacobian matrix. The dimension  $\dim N(J)$  of the zero space represents the arbitrary range of the given terminal velocity of the super redundant manipulator. From the physical point of view, these motions in the joint space will not cause the motion of the end, which is the self motion of the super redundant manipulator [7].

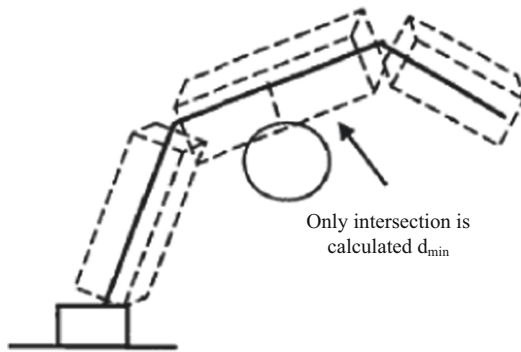
The above process completes the determination of the redundant space of the manipulator, which lays the foundation for the trajectory tracking planning of the manipulator.

### 2.4 The Obstacle Avoidance Index of Manipulator is Selected

The selection of obstacle avoidance indicators will directly affect the effectiveness of obstacle avoidance. Especially for dynamic obstacle avoidance, the random movement of obstacles, choosing a reasonable index that can accurately and quickly describe the position relationship between the obstacle and the manipulator is particularly important for ensuring the real-time nature of obstacle avoidance planning [8].

The traditional distance index needs to calculate the minimum distance from the obstacle to each member, and then take the minimum value. Take three members as an example,  $d_{\min} = \min\{d_1, d_2, d_3\}$ . When the bar is in some special configuration, the perpendicular foot of the minimum distance from the obstacle to the bar may fall on the extension line of the bar. The extension line does not belong to a part of the manipulator, so it is not accurate to calculate it by substituting the minimum distance. In this case, it is necessary to take multiple mark points on the bar and traverse the distance from each mark point to the obstacle. It can be seen that the traditional distance obstacle avoidance index is not accurate enough, but also needs to calculate multiple mark points, which costs a lot of time [9].

In view of the above problems, this paper preselects the minimum distance index  $H_{d-\min}$ , and its idea is: to establish corresponding bounding boxes for different manipulator configurations, use intersection test, eliminate other safety members before calculating the real-time distance, only consider the position relationship between the colliding members and obstacles, and calculate the distance value in the local coordinate system of the members, as shown in Fig. 3.



**Fig. 3.** Pre selected minimum distance Obstacle Avoidance Index

If there are multiple members intersecting in the intersection test, then  $H_{d-\min}$  takes the smallest of the intersection distances.  $H_{d-\min}$  index is used to describe the position relationship between the obstacle and the manipulator. Under the condition of avoiding the vertical foot from the obstacle to the bar falling on the extension line of the bar, the safety bar is removed, the traditional distance calculation is optimized, and the unnecessary amount of calculation is removed. In the process of obstacle avoidance, the larger the  $H_{d-\min}$  is, the better. It means that the farther away from the obstacle, the safer the manipulator is.

## 2.5 Trajectory Tracking Planning for Obstacle Avoidance of Manipulator

After obtaining the real-time minimum distance  $d_{\min}$  based on the above calculation, the trajectory tracking planning is carried out using the escape velocity dynamic

obstacle avoidance algorithm. Equation (4) provides the joint speed required to meet the terminal motion requirements, but if there are obstacles in the working space of the robot arm, one or more links may be too far away from the obstacle when the robot arm is in a certain position. Near or colliding with obstacles. To this end, this article first sets a “threshold” distance. During the movement of the manipulator arm, if the distance between each connecting rod and the obstacle is greater than the “threshold” distance, the position and posture of the manipulator arm can be obtained without changing the minimum norm solution (pseudo-inverse solution). On the contrary, the null space vector should be used to adjust the posture of the manipulator without changing the terminal posture, so as to meet the condition of no obstacles.

If the mechanical arm intersects with the obstacle at a certain time in the process of motion, the collision point between the link and the obstacle is  $x_o$ . In order to make the super redundant manipulator avoid obstacles, zero space vector should be selected, which is the closest point  $x_o$  and has the velocity component  $\dot{x}_o$  far away from the obstacles, so as to prevent the collision between the obstacles and the manipulator. Therefore, there are two basic requirements for obstacle avoidance trajectory tracking planning of super redundant manipulator: the first is to meet the terminal motion requirements, and the second is to avoid moving obstacles [10, 11], namely:

$$\begin{cases} J\dot{\theta} = \dot{X} \\ J_o\dot{\theta} = \dot{x}_o \end{cases} \quad (7)$$

In formula (7),  $\dot{x}_o$  is the set speed to avoid obstacles, and  $J_o$  is the Jacobian matrix of the collision point. Substitute formula (4) into formula (7), and increase the minimum distance  $d_{\min}$  from the obstacle avoidance index  $H_{d-\min}$ .

In order to reduce the influence of rank changes and maintain the continuity of joint speed, the obstacle avoidance gain  $a_n$  and the escape speed  $a_o$  are introduced. The dynamic obstacle avoidance inverse kinematics of the escape speed is expressed as:

$$\dot{\theta} = J^+ \dot{X} + a_n [J_o(I - J^+ J)]^+ (a_o \dot{x}_o - J_o J^+ \dot{X}) \quad (8)$$

In summary, through the introduction of VR technology, the tracking planning of the obstacle avoidance trajectory of the hyper-redundant manipulator is realized, which provides effective support for the application and development of the hyper-redundant manipulator.

### 3 Application Performance Analysis

#### 3.1 Experimental Preparation

For the application performance of the obstacle avoidance trajectory tracking planning method of the super-redundant manipulator based on VR technology, a simulation experiment is designed on the ground experimental system of the super-redundant manipulator based on the air-floating platform.

The ground experimental system of the hyper-redundant manipulator is an important part of the development process of the hyper-redundant manipulator. The ground experimental system based on the hyper-redundant manipulator can simulate a weightless environment, verify the validity of the theory related to the hyper-redundant manipulator, and test the performance and function of the manipulator. In the experimental system:

- 1) Air flotation platform: In order to simulate the microgravity environment in space, this experimental platform is based on the air flotation method, using the jet reaction force of the air bearing to support the seven-degree-of-freedom super-redundant manipulator on a flat and smooth marble platform. Counteract the effects of gravity. The air pump is used as the air source to supply air for each air foot to support the robotic arm to simulate a microgravity environment;
- 2) Central controller: coordinate and control the motion of the robotic arm, send joint commands to each joint controller, and the central controller and each joint controller communicate based on the 1553B bus protocol;
- 3) Ground inspection equipment: Located between the central controller and the simulation notebook, it plays the role of data forwarding, storage, viewing and management. The communication between the central controller and the ground inspection equipment is based on the 1553B bus protocol. The ground inspection equipment and the simulation notebook Communication is based on UDP protocol;
- 4) Simulation Notebook: realize VR technology, simulation and task planning function. At the same time, it communicates with the operator station to collect the operation status and send the sensor data to the operator station, which is based on UDP protocol;
- 5) Console: through the console, the operator sends control commands to check the status of the manipulator, which realizes the good interaction between the operator and the manipulator.

Due to the limitation of the experimental platform, it is difficult to realize the simultaneous motion of seven joints of the 7-DOF super redundant manipulator. The axes of joints 3, 4 and 5 of the 7-DOF super redundant manipulator are parallel to each other, so when joints 1, 2, 6 and 7 are fixed and only the end position of the manipulator is controlled, it is still a super redundant manipulator control problem. It can be used to verify the inverse kinematics and Obstacle Avoidance Trajectory Tracking planning of the super redundant manipulator, but the amount of calculation is less than that of the 7-DOF super redundant manipulator. In the experiment, keeping the angles of joints 1, 2, 6 and 7 at  $0^\circ$ ,  $-180^\circ$ ,  $0^\circ$  and  $90^\circ$ , joints 3, 4 and 5 move, which is equivalent to the motion of planar three-bar super redundant manipulator in plane  $X - Z$ .

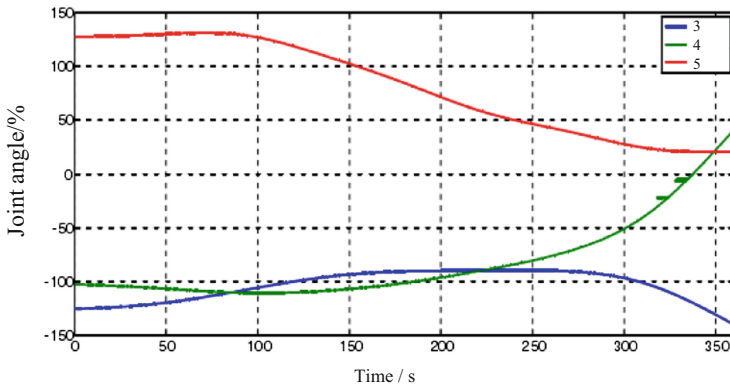
Although the 3, 4, and 5 joints of the hyper-redundant manipulator can move within the range of  $[-270^\circ, 270^\circ]$ , due to the limited area of the air bearing table and the restriction of the air foot support mechanism attached to the manipulator, The range of motion of the robotic arm on the experimental platform is limited. In the experiment, the range of motion of joint 3  $[-165^\circ, -90^\circ]$ , the range of motion of joint 4  $[-120^\circ, 120^\circ]$ , the range of motion of joint 5  $[20^\circ, 165^\circ]$ , the joints of joints 3, 4, and 5 The speed cannot exceed  $2^\circ/s$ , and the absolute position accuracy of the end is less than or equal to 10mm.



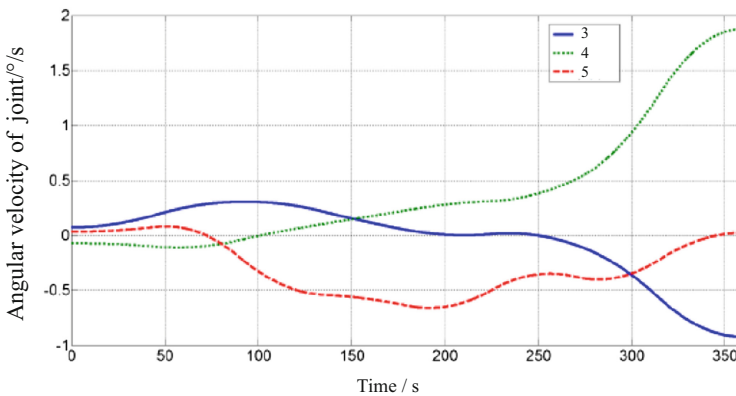
### 3.2 Analysis of Results

Task description: plan a collision free trajectory from the initial configuration to the target configuration, and make the trajectory cost of the end of the manipulator as small as possible. The initial configuration  $[-136.93^\circ, -74.31^\circ$  and  $24.57^\circ]$  and the target configuration  $[-139.92^\circ, 40.68^\circ$  and  $20.57^\circ]$ .

Through the simulation experiment, the change curve of joint angle and velocity in the process of obstacle avoidance movement of super redundant manipulator is obtained, as shown in Fig. 4.



(1) Joint angle curve



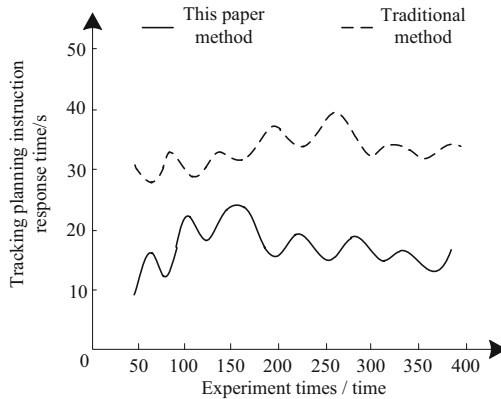
(2) Joint angular velocity curve

**Fig. 4.** Curve of joint angle and speed change during obstacle avoidance movement of the robotic arm

Can be seen from the Fig. 4, 3, 4, 5 joints Angle and angular velocity are within the scope of the constraints, and a smooth curve, can be used for the control of mechanical arm, arm successfully avoid the obstacles and reach the target configuration, today's

hyper redundant manipulator trajectory tracking of obstacle avoidance planning requirements, fully demonstrates the effectiveness of the proposed design method.

On this basis, the traditional trajectory tracking planning method based on machine vision technology is compared and further verified from the perspective of timeliness of response of tracking planning instructions. The results are shown in Fig. 5.



**Fig. 5.** Comparison of response timeliness of tracking planning instructions by different methods

According to Fig. 5, with the increase of verification times, the response time of tracking planning instructions of different methods is also changing constantly. However, it is obvious that the time of the proposed method is lower than that of the traditional method. It can be seen that the method presented in this paper can realize trajectory tracking planning more quickly.

## 4 Conclusion

In this study, VR technology was applied to the obstacle avoidance trajectory tracking planning of hyperredundant manipulator. In the design process, the hyperredundant manipulator is taken as the research object, the virtual model of the hyperredundant manipulator is constructed by using VR technology, and the redundant space of the manipulator is determined. Then, the pre-selected minimum distance index was selected as the obstacle avoidance index of the manipulator. According to the calculated real-time minimum distance, the inverse kinematics equation of the manipulator was obtained by using the escape velocity dynamic obstacle avoidance algorithm, so as to realize the trajectory tracking planning. The experimental results show that this method can effectively meet the requirements of obstacle avoidance trajectory tracking planning for hyperredundant manipulators and is suitable for extensive application. In the following research, the proposed method will be further optimized from the perspective of improving the control accuracy.

**Fund Projects.** Supported by the National Key Research and Development Program of China (No. 2019YFC1408003).

## References

1. Aijuan, L., Wanzhong, Z., Xibo, W., et al.: ACT-R cognitive model based trajectory planning method study for electric vehicle's active obstacle avoidance system. *Energies* **11** (1), 75 (2018)
2. Mu, Z., Liu, T., Xu, W., et al.: A Hybrid Obstacle-avoidance method of spatial hyper-redundant manipulators for servicing in confined space. *Robotica* **37**(6), 998–1019 (2019)
3. Yue, M., Wu, X., Guo, L., et al.: Quintic polynomial-based obstacle avoidance trajectory planning and tracking control framework for tractor-trailer system. *Int. J. Control Autom. Syst.* **17**(10), 2634–2646 (2019)
4. Hu, Z., Zhu, D., Cui, C., et al.: Trajectory tracking and re-planning with model predictive control of autonomous underwater vehicles. *J. Navig.* **72**(2), 321–341 (2019)
5. Kano, H., Fujioka, H.: Spline trajectory planning for road-like path with piecewise linear boundaries allowing double corner points. *Sice J. Control Meas. Syst. Integrat.* **11**(6), 429–437 (2018)
6. Xiuxia, Y., Yi, Z., Weiwei, Z.: Obstacle avoidance method of three-dimensional obstacle spherical cap. *J. Syst. Eng. Electron.* **29**(05), 182–192 (2018)
7. Chiaravalli, D., Califano, F., Biagiotti, L., et al.: Physical-consistent behavior embodied in B-spline curves for robot path planning. *IFAC-PapersOnLine* **51**(22), 306–311 (2018)
8. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
9. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. *Front. Genet.* **10**, 120 (2019)
10. Liu, S., Bai, W., Zeng, N., et al.: A fast fractal based compression for MRI images. *IEEE Access* **7**, 62412–62420 (2019)
11. Wu, Q., Zhang, C., Zhang, M., et al.: A modified comprehensive learning particle swarm optimizer and its application in cylindricity error evaluation problem. *Int. J. Performabil. Eng.* **15**(3), 2553 (2019)



# High Voltage Cable Shield Voltage Monitoring Method Considering Coil Flux

Qiu-jiao Huang<sup>(✉)</sup> and Dan-kang He

Guangxi Modern Polytechnic College, Hechi 547000, China  
huangqiuqiao3259@163.com

**Abstract.** The traditional high-voltage cable shielding layer voltage monitoring method adopts the phase comparison method, which is easy to be interfered by external electrostatic induction, magnetostatic induction and other factors, resulting in low monitoring accuracy and efficiency. In view of the above problems, this paper studies the high-voltage cable shielding layer voltage monitoring method considering coil magnetic flux. After calculating the transfer impedance of cable shielding layer, the calculation model of high voltage cable magnetic flux mutual inductance parameters is constructed. The parameters such as induced voltage of cable shielding layer are determined, and the collected voltage signal is processed by morphology. Using Duffing chaos detection principle, the shielding layer voltage monitoring results are obtained. The experimental results show that the proposed method has high monitoring efficiency, and the minimum effective monitoring rate is 94.3%.

**Keywords:** Coil flux · High voltage cable · Cable shielding layer · Voltage monitoring

## 1 Introduction

Compared with fossil energy, electric energy has the advantages of easy control and convenient transmission. After the outbreak of the second industrial revolution, electric energy has gradually become the preferred energy for many industrial sectors. High voltage cable is an important part of power transmission, high-voltage cable transmission power, causing a greater transmission burden to the cable. Therefore, the high-voltage cable is composed of multi-layer structure, and the shielding layer is a layer of metal material on the surface of the cable and connector, which can ensure the normal operation of the cable [1–3]. In general, the shielding layer of high-voltage cable is used to balance the unbalanced current generated by three-phase alternating current, meet the short-circuit current requirements of single-phase grounding fault when the cable is running, and protect the high-voltage cable from the effect of electric force under the changing unbalanced magnetic field, so that the power cable terminal is under the effect of mechanical force caused by electric force for a long time, especially for the stress area of the cable terminal, it is easier to cause damage, leading to electrical insulation damage and shortening the service life of the cable terminal. Monitoring the grounding effect of high-voltage cable shielding layer is an important way to timely pay attention to the safety of high-voltage cable operation. By collecting, transmitting and

analyzing the current and voltage signals of cable shielding layer, the state of shielding layer can be obtained in real time to achieve the purpose of safety protection. The traditional high voltage cable shielding layer monitoring method uses hardware data collector to collect the shielding layer status data in the section, and obtains the monitoring results under manual processing. This traditional monitoring method has high requirements for hardware equipment and technical personnel, and it is difficult to ensure more accurate monitoring effect. The online monitoring and signal transmission technology provides the basis and convenience for the intelligent development of power system [4].

In the power system, the usual test method to judge whether the cable insulation shielding layer is good or bad is to apply over-voltage to the tested cable insulation. The common methods include DC withstand voltage and AC withstand voltage to test whether the cable insulation withstand voltage passes or not, and the DC leakage current can also be detected when DC withstand voltage. However, the current intensity of high-voltage cable transmission will be affected by the grid load, and its data monitoring is more complex. Therefore, monitoring voltage is usually used to monitor the working state of high voltage cable shielding layer. The monitoring method based on the dielectric loss method compares the phase of the signal obtained on all monitoring cables with the voltage signal on the bus to monitor the dielectric loss factor of all cables [5]. This method can reflect the overall situation of the shielding layer, but the external interference factors such as serious electrostatic induction, magnetostatic induction and corona discharge with wide spectrum will lead to large monitoring deviation. The high voltage cable shielding layer voltage monitoring method based on LabVIEW virtual instrument for upper computer and lower computer makes use of the high measurement accuracy and powerful function of the data acquisition card provided by LabVIEW, but it is not conducive to the expansion and large-scale application of the monitoring method, and the price is expensive, which is not conducive to cost saving. Based on the above analysis, this paper will study the high voltage cable shield voltage monitoring method considering coil flux.

## **2 Voltage Monitoring Method of High-Voltage Cable Shielding Layer Considering Coil Magnetic Flux**

### **2.1 Calculation of Transfer Impedance of Cable Shielding Layer**

Whether in the time domain or the frequency domain to establish the electromagnetic field coupling model to the shielded cable, it is necessary to know the distribution parameters along the line and the transfer impedance of the shielding layer. Transmission line distribution parameters are the basis for establishing external cable coupling, while transfer impedance is the bridge connecting the internal and external loop couplings. The factors that affect the coupling are the inductance, capacitance, conductance and resistance per unit length, mainly the solution of distributed inductance and distributed capacitance. The solution to the distribution parameters is closely related to the selection of the reference conductor, including one of the transmission lines as the reference conductor, the shielding layer as the reference conductor, and the

ground as the reference conductor. The coupling of a complete shielded cable is divided into internal and external transmission systems. The external transmission system and the internal transmission system use the earth and the shielding layer as reference conductors respectively.

The transverse electromagnetic field structure and the corresponding propagation mode are the basic assumptions of the theoretical analysis of the transmission line. Even if the electromagnetic field changes with time, the voltage and current between the transmission line conductors can still be uniquely defined. The general simplified form of the equation of  $n + 1$  multi-conductor transmission line is as follows [6]:

$$\begin{aligned} \frac{\partial}{\partial z} V(z, t) &= -RI(z, t) - L \frac{\partial}{\partial z} I(z, t) \\ \frac{\partial}{\partial z} I(z, t) &= -GV(z, t) - C \frac{\partial}{\partial z} V(z, t) \end{aligned} \tag{1}$$

Among them,  $R$  and  $G$  are the impedance matrix per unit length and the conductance matrix per unit length respectively, which are ignored for lossless conductors.

$$L = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1n} \\ l_{21} & l_{22} & \cdots & l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \tag{2}$$

$$C = \begin{bmatrix} \sum_{k=1}^n c_{1k} & \cdots & -c_{1m} & \cdots & -c_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ -c_{m1} & \cdots & \sum_{k=1}^n c_{mk} & \cdots & -c_{mn} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ -c_{n1} & \cdots & -c_{nm} & \cdots & \sum_{k=1}^n c_{nk} \end{bmatrix} \tag{3}$$

The matrices  $L$  and  $C$  are the distributed inductance matrix and the distributed capacitance matrix, respectively.  $l_{ji}$  and  $l_{ij}$  are the self-inductance of the  $i$  conductor and the mutual inductance of the  $i$  conductor and the  $j$  conductor, respectively.  $c_{ii}$  and  $c_{ij}$  are the self-capacitance of the  $i$  conductor and the mutual capacitance of the  $i$  conductor and the  $j$  conductor. Because the mutual inductance and mutual capacitance between the two conductors are equal,  $l_{ji} = l_{ij}$ ,  $c_{ii} = c_{ij}$ , the inductance matrix and the capacitance matrix are all symmetrical.

First, analyze the basic sub-problem of magnetic flux generated by a wire with a uniformly distributed current  $I$ . For the transverse electromagnetic field, using the Ampere’s law loop, assuming that the circular magnetic field at the same distance from the center of the wire is of the same size, the area along the circumference of the

distance  $r$  is divided, and the magnetic flux per unit length can be obtained as follows [7]:

$$\int_l H dl = I$$

$$H_\phi = \frac{I}{2\pi r}$$

$$\psi = \frac{\mu I}{2\pi} \ln\left(\frac{r_2}{r_1}\right)$$
(4)

In the above formula,  $H$  is the magnetic flux;  $l$  is the current component;  $I$  is the wire current;  $H_\phi$  is the cross-sectional magnetic flux of the wire;  $r_1$  and  $r_2$  are the distance between the wire and the magnetic flux surface. In the same way, the entire magnetic flux wire can be obtained.

When the spacing of the high-voltage transmission line is wide, the dielectric insulating layer of the transmission line only slightly changes the capacitance distribution around the wire. If the surrounding medium is considered to be uniform, the inductance matrix and the capacitance matrix have the following important relationship [8]:

$$LC = CL = \mu\epsilon A$$
(5)

In the formula,  $\mu$  and  $\epsilon$  are the permeability and permittivity of the medium surrounding the conductor, and  $A$  is the identity matrix. Through the obtained unit distribution inductance matrix, the unit capacitance matrix can be calculated.

For shielded cables with the shielding layer as the reference conductor, the inner shielding layer is a multi-conductor transmission line. The radius of the shielding layer is  $r_{SH}$ , and the distance between the wire  $i$  and the wire  $j$  from the center of the shielding layer axis is the angle formed by  $d_i$  and  $d_j$  as  $\theta_{ij}$ . Similar to the ground plane as the reference conductor, replace the shielding layer with mirrored wires.

It can be obtained from inductance and mutual inductance as [9]:

$$l_{ii} = \frac{\mu_0}{2\pi} \ln\left(\frac{r_S^2 - d_i^2}{r_S r_i}\right)$$

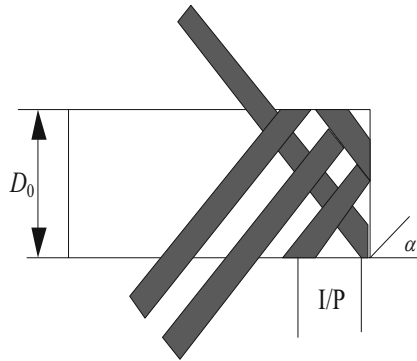
$$l_{ij} = \frac{\mu_0}{2\pi} \ln \left[ \frac{d_j}{r_S} \sqrt{\frac{(d_i d_j)^2 + r_S^4 - 2d_i d_j r_S^2 \cos \theta_{ij}}{(d_i d_j)^2 + r_S^4 - 2d_i d_j^3 \cos \theta_{ij}}} \right]$$
(6)

Find the distributed capacitance matrix through the relationship between  $L$  and  $C$ .

Braided shielding layer is widely used due to its good shielding effect. For high frequency electromagnetic interference, a braided shielding sheath with high coverage is generally used. However, as the frequency increases, electromagnetic energy will be coupled into the inside of the sheath through the woven mesh, and it will also cause certain interference to the cables inside the sheath. Therefore, the transfer impedance is

used to indirectly measure the shielding performance of the shielding layer. The smaller the transfer impedance, the better the shielding performance, and vice versa.

The braided layer of the braided shielding sheath is generally copper tinned material. The weaving bundles of opposite directions are cross-knitted in and out to form many tiny diamond-shaped holes, which saves materials and prevents the weaving coverage rate from reaching 100%. The method of solving the transfer impedance of the shielded cable layer is to analyze the coupling mechanism of the external electromagnetic field to the braided shielding layer, and then further derive the calculation formula. The most basic braiding parameters are as follows: (1) the inner diameter of the shielding layer  $D_0$ ; (2) the number of braided cross strands per unit length  $p$ ; (3) the number of braided strands  $c$ ; (4) the number of wires per share  $N$ ; 5) The diameter  $d$  of each braided thin wire (Fig. 1).



**Fig. 1.** Side development view of braided shielding layer

Analyze the coupling mechanism of the electromagnetic field on the braid. When the external time-varying electromagnetic field is irradiated on the braided layer with incomplete coverage, scattering, transmission and diffraction will occur on the shielding layer.

According to the theoretical analysis of electromagnetic field, the transfer impedance formula of braided cable is as follows [10]:

$$Z_t = Z_d + (M_h + M_b) \quad (7)$$

In the formula,  $Z_d$  means that the transient electromagnetic field will produce scattering characteristics when radiating onto the braided layer, which is a good representation of its low-frequency characteristics, and it will drop rapidly as the frequency increases.  $M_h$  is the small hole inductance caused by the transmission of the magnetic field through the mesh of the braid.  $M_b$  is the braided inductance generated by the cross braiding of braided bundles and the magnetic lines of force in the gap between the braided bundles. These three parts together constitute the transfer impedance of the braided shielding layer.



Table 1 below shows the relationship between the number of braided strands and coverage [11].

**Table 1.** The relationship between the number of braided strands and coverage

Number of shares N	Inner diameter of shielding layer $D_0$	Braided corner $l^\circ$	$d$	C	Thickness of braid coverage %
6	4.8	20	0.15	16	71.33
6	4.8	20	0.15	18	77.21
6	4.8	20	0.15	20	82.42
6	4.8	20	0.15	22	86.95
6	4.8	20	0.15	24	90.81

At present, the most used and theoretically accurate calculation formula for scattering impedance is derived from the research of Vance, which characterizes the low-frequency characteristics of electromagnetic field scattering to the braid. After calculating the transfer impedance of the cable shielding layer according to the above process, a calculation model for the magnetic flux mutual inductance parameters of the high-voltage cable is constructed.

## 2.2 Construction of Calculation Model for Magnetic Flux Mutual Inductance Parameters of High-Voltage Cables

Unbalanced three-phase currents, unequal lengths of cross-interconnected segments, asymmetrical geometric arrangement, etc. will all cause the sheath current to increase. The sheath impedance is greater than the conductor impedance, and the additional temperature rise caused by the heating of the sheath current will reduce the current carrying capacity of the cable and increase the temperature of the cable. When the outer sheath is damaged, it may even cause thermal breakdown of the cable. Therefore, the magnetic flux mutual inductance parameters of high-voltage cables are calculated by constructing a model.

The defined average diameter  $d_s$  (unit: m) of the sheath of the shielding layer is shown in the following formula.

$$d_s = \frac{d_{se} + d_{si}}{2} \quad (8)$$

Among them,  $d_{se}$  is the outer diameter of the corrugated sheath, and  $d_{si}$  is the inner diameter of the corrugated sheath. To calculate the mutual inductance of the sheath, the upper and lower limits of the integral need to be determined, that is, the thickness  $t_s$  of the sheath needs to be considered. For the convenience of comparison, this paper uses the geometric mean radius (equivalent radius) to calculate the mutual inductance of the cable, and the approximate solution obtained is consistent with the equivalent diameter

method. Assuming the equivalent radius, the protective layer can be equivalent to a hollow cylinder with an inner radius of  $r_s - 0.5t_s$  and an outer radius of  $r_s + 0.5t_s$ .

According to Biot-Savart's law, the magnetic induction intensity  $B$  (unit: T) generated by the core current  $I$  at a distance from the wire  $r$  is:

$$B = \frac{\mu_0}{4\pi} \int_{\theta_1}^{\theta_2} \frac{I \sin \theta d\theta}{r} \tag{9}$$

Among them,  $\theta_1$  and  $\theta_2$  respectively fix the angle between the first and last ends of the straight wire and the line at  $\rho$  and the current element. When the core is long enough,  $\theta_2 = \theta_1 \approx \pi/2$ .

According to Gauss's theorem, the magnetic flux  $\Phi$  (unit: Wb) passing through the sheath is equal to the area fraction of the magnetic induction intensity vector  $B$  generated by the core current to the section  $S$  of the sheath:

$$\Phi = \iint_S B \cdot dS \tag{10}$$

The mutual inductance  $M$  (unit: H/m) between the core and the sheath of the high-voltage cable of length  $z$  can be obtained according to Faraday's law of electromagnetic induction [12, 13]:

$$M = \frac{\Phi}{I} = \frac{\mu_0}{2\pi} \iint_S \frac{z}{sr} dS \tag{11}$$

Among them,  $S$  is the position between the outer edge of the wire core and the center line of the metal sheath.

Currently, there are three manufacturing processes for corrugated sheaths: extrusion, hydrogen arc welding, and embossing. The embossing shape of the sheath mainly has three types: threaded, ring-shaped and smooth, of which the threaded shape is the most common. The inner and outer surfaces of the thread-shaped sheath are curved helical surfaces formed by the spiral motion of a curve (press roller or pressing wheel), and the thickness of the sheath remains unchanged. The threaded cable moves at a uniform speed along the axis, and the pressure roller or pressure wheel continues to be extruded. The movement pitch distance of the ring-shaped cable is squeezed by the pressure roller or the pressure wheel once. The sheath of the smooth sheathed cable does not need to be squeezed, and can be equivalent to a coaxial cylinder.

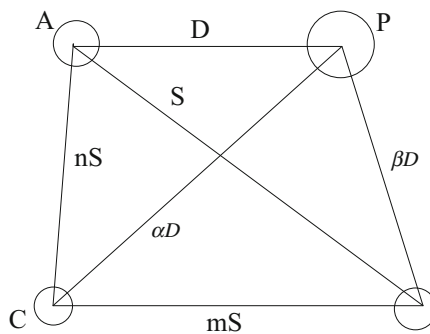
For corrugated (threaded and ring-shaped) metal sheathed high-voltage cables, the cross-sectional geometry of the magnetic line of force and the sheath is different from the smooth sheath. When the corrugated metal sheath formula (11) is used to calculate the mutual inductance of cables, the geometric shape of the sheath corrugation needs to be considered to determine the integral boundary conditions and solve the analytical solution of mutual inductance. In this paper, the parameter equations of the ring-shaped and spiral-shaped corrugated sheaths are established as the boundary conditions of Eq. (11), and the analytical solution of the mutual inductance of the ring-shaped and threaded sheathed cables is derived. Through the projection of the shielding layer, the

corresponding structural parameter equation of the shielding layer is established. According to the structure parameter equation of the shielding layer, the calculation model of the magnetic flux of different cable shielding layers in the electric field can be obtained. After the above-mentioned cable magnetic flux mutual inductance parameter model is established, the induced voltage of the cable shielding layer is calculated.

### 2.3 Calculation of Induced Voltage on Cable Shield

For single-core cables, the induced voltage of the metal sheath is completely different from that of three-core cables. The induced voltage on the metal sheath of a single-core cable depends on the load current of the cable and the arrangement and line length of the cables in the same circuit, as well as the arrangement and distance of the surrounding circuits. The relationship between the conductor and the metal sheath of a single-core cable can be regarded as the primary winding and secondary winding of a transformer. When the cable wire passes current, magnetic flux is generated around it. The magnetic flux is not only connected to the core loop, but also connected to the metal sheath of the cable, and induced voltages are generated on the core and the metal sheath respectively. Since each core has a dedicated metal protective layer, the magnetic flux generated by the load current or short-circuit current is interlinked with the metal protective layer, so there is always an induced voltage on the metal protective layer. The value of this induced voltage is related to the cross-section of the core, the distance between the cables and the current and the length of the cable. Especially when the cable is very long, the induced voltage on the sheath can reach a higher value.

In Fig. 2, *p* represents the metal sheath of the single-core cable, which can be regarded as a conductor parallel to the three-phase cores A, B, and C. The distance between the centers of the four conductors is expressed as a ratio, that is, the center distances between the cores AB, BC, and CA are *S*, *mS*, and *nS*; the center distances between the conductor *p* and the cores A, B, and C are respectively *D*,  $\beta D$ , and  $\alpha D.$



**Fig. 2.** Representation of the center distance between the metal sheaths of a single-core cable

According to the above analysis, it can be seen that there is an induced electric potential in the metal sheath, and the metal sheath connects the cross interconnection to

protect the ground to form a loop with the earth. Circulating current will be generated when the three-phase grounding is unbalanced due to grounding methods, grounding defects, and grounding faults caused by metal sheath defects. Because this current is generated by the induced potential of the metal sheath acting on the ground loop, it is called this current It is grounding induction circulating current.

Based on the above capacitance parameters, the power cable in operation will generate capacitance current. When one end of the metal shield is grounded, the ground capacitance current and leakage current of the cable have a path. Under normal circumstances, the capacitor current is very small.

When the single-core cable is laid straight and parallel, the equivalent circuit diagram of the metal sheath normally grounded is shown in Fig. 3.

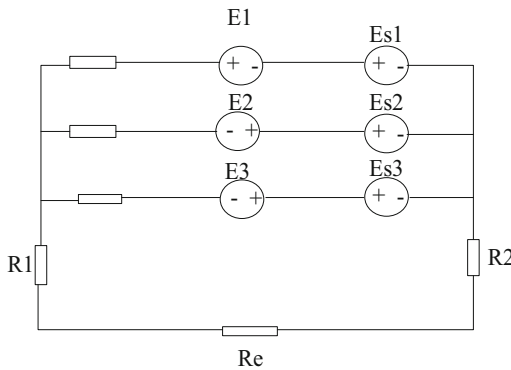


Fig. 3. Equivalent circuit diagram of metal protective layer grounding

$E_1$ ,  $E_2$ , and  $E_3$  are the unit induced potentials of the metal shielding layer of the three-phase cable due to the main current of the respective cores;  $E_{s1}$ ,  $E_{s2}$ , and  $E_{s3}$  are the unit inductions of the metal shielding layer of the two-phase cable to the other phase of the shielding layer. Electric potential;  $R_1$ ,  $R_2$  are the grounding resistance of the cable metal shielding layer;  $R_e$  is the earth resistance (including the leakage resistance of the cable to the earth  $R_g$ );  $R$  is the cable metal shielding layer resistance;  $X$  is the self-inductance of the cable metal shielding layer;  $X_1$  is the unit length The mutual inductance of the metal shielding layer of the middle-phase and side-phase cables;  $X_2$  is the mutual inductance of the metal shielding layer of the side-phase and side-phase cables per unit length.

Through the above derivation, the general calculation formula of the induced voltage of each phase of the cable is obtained, but it is far from enough for the study of the sheath induced voltage, because the above formula does not show the relationship between the induced voltage and the grounding resistance, whether the core transposes or not, and the number of cross interconnection points. When the length of the line exceeds 1000 m, the cable cross interconnection grounding mode is generally adopted, and this mode can also be subdivided into the metal sheath cross interconnection grounding mode and the metal sheath cross interconnection mode, and the wire core is

transposed at the same time. For the convenience of description, these two grounding modes are referred to as semi transposition grounding mode and full transposition grounding mode respectively. According to the relevant simulation research, whether in the horizontal arrangement or the zigzag arrangement, the sheath induced voltage is negatively correlated with the number of cross interconnection points, and is always positively correlated with the distance. However, in the actual cable laying, the smaller the distance is, the better. Therefore, the cable diameter, construction conditions and induced voltage limit should be considered comprehensively, choose the best distance.

When the cables are arranged horizontally, try to choose the full transposition grounding method, because the full transposition is compared with the half transposition, the sheath induced voltage value is significantly reduced, and the overall induced voltage of the line is in a balanced state.

When the cables are arranged in a zigzag pattern, the full transposition and semi transposition grounding methods have almost the same induced voltage curve. Because the full transposition has greater construction difficulty in the actual engineering construction, the semi transposition grounding method can be selected when the induced voltage requirement is not high.

When the cable line is grounded through resistance at both ends, there is a certain relationship between the sheath induced voltage and the grounding resistance, and the voltage increases with the increase of resistance. When the cable line is in the cross connected grounding state, the grounding resistance has little effect on the induced voltage. After determining the quantization process of the above factors affecting the voltage of the cable shielding layer, the voltage signal collected from the voltage data is processed.

## 2.4 Shield Voltage Signal Processing

After the collected high-voltage cable shielding layer voltage signal undergoes preliminary processing by the hardware part, the signal image is obtained, and the shielding layer voltage signal is processed according to the morphological theory.

The working principle of mathematical morphology is to use structural elements to do morphological operations on the areas that need filtering to filter and analyze the graphics and signals. The parameters of structural elements such as length, shape and the selected morphological operations determine the final filtering results. Because most of the images in real life and the main research are not binary images, but gray images with pixel gray values greater than two, the gray morphology is also used after the promotion of binary morphology. The signal collected from the power system is one-dimensional signal. Aiming at the gray-scale morphological transformation in the case of one-dimensional discrete, there is a new definition of corrosion and expansion operation. The maximum and minimum operation can be used to replace the intersection and union operation in the corrosion and expansion operation defined in the binary image.

Assuming that the input signal  $f(n)$  is a discrete function defined on  $F = (0, 1, \dots, N - 1)$ , and the structural element sequence  $g(n)$  is defined as a discrete function on  $G = (0, 1, \dots, M - 1)$ , and  $N \geq M$ , the newly defined corrosion expansion operation is as follows:

$$\begin{aligned}(f \ominus g) &= \min(f(n+m) - g(m)) \\ (f \oplus g) &= \max(f(n+m) + g(m))\end{aligned}\tag{12}$$

The corrosion using the new definition of formula (12) is actually an operation to take a minimum value, while the expansion using the new definition of formula (12) is actually an operation to take a maximum value. In the same way, more complex morphological operations can be derived from the two basic operations of grayscale corrosion and expansion operations. Mathematical morphology operation is simple, mainly composed of four basic operations: expansion, corrosion, opening operation and closing operation, which can quickly process and analyze the input signal.

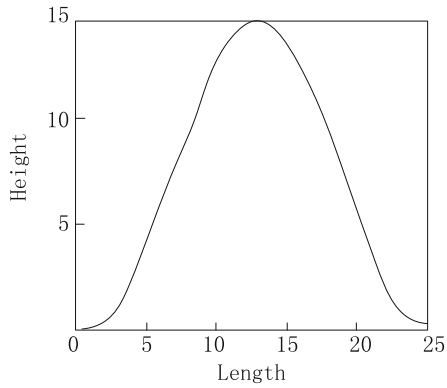
When using a mathematical morphological filter to filter a sampled signal, in addition to the selected mathematical morphology algorithm, the other factor that affects the filtering effect is the selection of structural elements. The selection of structural element parameters mainly includes two aspects: one is the shape of structural elements. Common shapes include straight lines, semicircles, cosines, triangles, etc. The second is the size of the corresponding structural element shape. Structural elements of different shapes and sizes will produce different filtering effects even if they are combined with the same filtering algorithm.

According to the research of experts and scholars at home and abroad in recent decades, the semicircular structural element has the best performance in filtering white noise, the triangular structural element has the best performance in filtering positive and negative impulse noise, and the linear structural element has strong comprehensive denoising ability, which is suitable for processing signals with high transient component content.

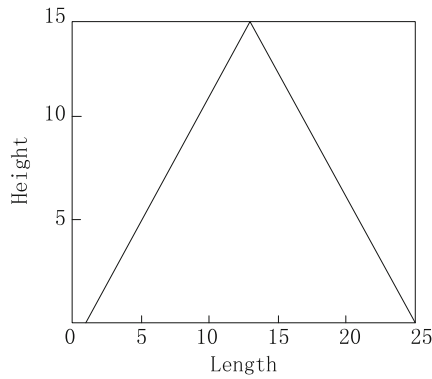
In engineering practice, the collected signal is more complex, including many different kinds of noise, so the more structural elements, the better filtering effect will be. But on the other hand, the amount of calculation and the delay are proportional to the shape, size and complexity of structural elements. In order to give consideration to both the filtering effect and the amount of calculation, the following two principles should be followed when designing the parameters of the filter: firstly, when selecting the size of the structural elements, the selected size should not be too long, and should be as short as possible to reduce the workload. Generally, the length of the selected structural elements is between 1/50 and 1/20 of the signal cycle length; Secondly, when selecting the shape of structural elements, we should select the structural elements similar to the shape of the signal to be filtered.

The working environment of power grid is complex, and the sampled power signals are often mixed with many different types of noise. When morphological filters use different shapes of structural elements, the filtering performance will be different. Each shape of morphological filter has its own characteristics, which is suitable for filtering different characteristics of noise. In order to remove all kinds of interference signals more effectively, this paper uses a composite morphological filter, which is composed of basic structural elements with different shapes, and uses the new structural elements to filter and analyze the signals. In this paper, the triangle structure element which has the best performance in filtering positive and negative impulse noise and the sine structure element which is most similar to the waveform of power signal are combined

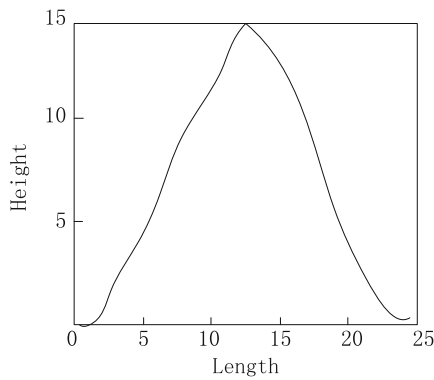
to form a new morphological filter, which can also play a good filtering effect on the signal disturbed by complex noise.



a) Structural elements with a sine shape



b) Structural elements shaped as triangles



c) Composite structural elements

**Fig. 4.** Working principle of complex morphological filter

Figure 4a) is a sine-shaped structural element, Fig. 4b) is a triangular-shaped structural element, and Fig. 4c) is a new structural element composed of the above two structural elements. After processing the voltage signal of the shielding layer by using the compound morphological filter of the above structure, the chaos theory is used to realize the voltage monitoring of the shielding layer.

## 2.5 Realize Shield Voltage Monitoring

After continuous research on the Duffing equation, it is found that the Duffing system can express various properties through its dynamic behavior. The nonlinear dynamic system expressed by the Duffing system has a variety of nonlinear dynamic characteristics, including various complex dynamics such as oscillation, chaos, and bifurcation. This has also become one of the commonly used models for chaos research. The buffing chaotic oscillator equation is a nonlinear vibration system with nonlinear restoring force, and the form is as follows:

$$\ddot{x}(t) + k\dot{x}(t) - x(t) + x^3(t) = a \cos \omega t \quad (13)$$

Among them,  $k$  is the damping ratio;  $\omega$  is the angular frequency of the periodic external force;  $a$  is the amplitude of the periodic external force. According to related research:

- 1) The amplitude of periodic external force affects the trajectory and mode of chaotic motion. Under the same periodic external force frequency, when the amplitude of the periodic external force changes, the dynamic behavior will also change, and the specific manifestation is that the phase trajectory produced by the chaotic motion will also change.
- 2) The damping coefficient of the system will also affect the threshold of chaotic motion. At the same frequency, when the damping ratio increases, the periodic external force amplitude  $a$  that causes chaotic motion will also increase, and its frequency range will also increase. But if the damping ratio is too large, it will cause the chaotic motion to change to a periodic motion.
- 3) Through the analysis of the phase trajectory changes of the system, it can be explained that the dynamic behavior in the chaotic system is very sensitive to the initial parameters. Therefore, according to the characteristics of the system's dynamic behavior from the chaotic critical state to the phase trajectory of the large-scale periodic state, the detection of the weak signal to be measured is carried out. This is also a practical method of applying the Duffing chaotic system to the theory of detecting weak periodic signals.

The model of using Duffing chaos to detect periodic signals is as follows:

$$\ddot{x} - x^3 + x^5 + k\dot{x} = \gamma \cos \omega \quad (14)$$



Among them,  $k$  is the damping ratio;  $-x^3 + x^5$  is the nonlinear restoring force of the system;  $\gamma \cos \omega t$  is the applied periodic force of the system. The basic idea of using this Duffing chaotic oscillator for periodic signal detection is: first adjust the system parameter  $\gamma$  to reach the critical point  $\gamma_0$ , so that the system is in the state of the boundary between the chaotic state and the large periodic motion state, and then use this signal to be detected as the period of the system's periodic external force. Perturbation is based on the phase change between the chaotic state and the ordered state of the system, and then determines whether the signal to be measured appears, and then adjusts the relevant parameters to achieve the purpose of signal measurement. The motion state of the system can be characterized by the maximum Lyapunov exponent  $\lambda$  of the system. When  $\lambda$  is greater than 0, the system must be in a chaotic state, and the greater the value of  $\lambda$ , the deeper the chaotic depth; when  $\lambda$  is less than 0, the system belongs to Ordered system, that is, periodic system or quasi-periodic system.

According to the above-mentioned Duffing chaotic oscillator to discriminate the system state, the voltage state and value of the shielding layer in the current chaotic period can be obtained accordingly. Comparing the output result of the chaotic system with the voltage state of the shielding layer during normal operation, the monitoring result of the shielding layer voltage of the high-voltage cable is obtained. The above has completed the research on the voltage monitoring method of the high-voltage cable shielding layer considering the coil magnetic flux.

### 3 Experimental Study

This section will verify the feasibility of the high-voltage cable shielding voltage monitoring method that considers the coil magnetic flux.

#### 3.1 Experiment Content

The experiment adopts the form of comparison, using two traditional voltage monitoring methods as Method 1 and Method 2, to verify the feasibility and reliability of the method from the perspective of monitoring error and monitoring response time of the monitoring method. Three monitoring methods are used to monitor the shielding layer voltage of cables in the same power grid, and the actual value is compared with the monitoring value to obtain an effective monitoring rate; the time when the monitoring data change occurs is recorded as the monitoring response time.

#### 3.2 Experimental Result

The experimental results are shown in the following table. Analyze the data in the table and draw the corresponding experimental conclusions (Table 2).

**Table 2.** Experimental data

Cable number	Method 1		Method 2		Method of this article	
	Effective monitoring rate/%	Monitoring reaction time/ms	Effective monitoring rate/%	Monitoring reaction time/ms	Effective monitoring rate/%	Monitoring reaction time/ms
1	82.2	51.9	88.5	50.3	95.3	29.1
2	81.4	50.5	88.1	50.7	94.4	29.8
3	78.8	46.7	86.8	49.6	95.7	27.9
4	84.6	53.2	86.8	49.8	94.5	27.8
5	85.0	47.4	88.1	49.6	94.8	28.6
6	83.2	51.2	87.7	50.4	95.3	28.2
7	80.7	49.7	88.5	49.5	95.2	28.1
8	79.1	53.3	87.7	49.1	95.1	27.9
9	84.3	49.1	88.4	48.9	95.6	28.6
10	81.7	42.6	86.6	50.2	95.9	29.7

Analysis of the data in the above table shows that the effective monitoring rate of method 1 is lower than 85%, the effective monitoring rate of method 2 is lower than 88.6%, and the lowest effective monitoring rate of this method is 94.3%, indicating that the monitoring accuracy of this method is higher. In terms of reaction time, the average reaction time of the method in this paper is 28.57 ms, the average reaction time of method 1 is 49.56 ms, and the average reaction time of method 2 is 49.81 ms. The method in this paper is lower, indicating that the method is more efficient. That is to say, the high-voltage cable shielding voltage monitoring method that considers the coil magnetic flux studied in this paper has the advantages of high monitoring accuracy and fast speed.

## 4 Concluding Remarks

The metal shield grounding of high voltage power cable can effectively ensure that the lines and electrical equipment are not damaged and maintain the stable operation of power system. However, the current application of high-voltage power cable metal shield grounding method has not formed a unified standard. If the correct grounding method is not adopted, it will lead to power accidents, which will not only endanger people's lives, but also cause serious disasters to enterprises. The voltage of the shielding layer of high voltage cable can effectively ensure the normal work of the shielding layer. This paper studies the method of monitoring the shielding layer voltage of high voltage cable considering the coil flux. Based on the calculation of transmission impedance of cable shielding layer, the calculation model of magnetic flux mutual inductance parameters of high voltage cable is established. The induced voltage and other parameters of the cable shielding layer are determined, and the collected voltage signal is processed by morphology. Using Duffing chaos detection principle, the monitoring results of shielding layer voltage are obtained. The experimental results show that the voltage monitoring method has the advantages of high precision and strong stability.

**Fund Projects.** 1. 2020 Guangxi University Young and middle-aged teachers' scientific research basic ability promotion project: Research on classroom teaching dynamic evaluation system based on mobile terminal (2020ky45011).

2. 2016 Guangxi Vocational Education Teaching Reform Project: Study on the Connection of Project Design between Specialized Courses in "Project Teaching" of Applied Electronic Technology Major in Higher Vocational Education, (GXGZJG2016B003)

## References

1. Liu, Y., Xia, X., Li, M., et al.: Research on online monitoring system based on locus method of HV power cable. *J. Electric Power Sci. Technol.* **34**(03), 202–210 (2019)
2. Wei, D., Yuguang, S., Daizong, T., et al.: Online monitoring method for rotor demagnetization fault of permanent magnet synchronous machine based on new type of search coil. *Electric Power Autom. Equipment* **40**(06), 218–227 (2020)
3. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
4. Zhao, W., Xia, X., Liu, Y., et al.: Research on online monitoring for HV cable running locus. In: Proceedings of the CSU-EPSCA, vol. 31(09), pp. 137–143 (2019)
5. Qi, H., Zhang, B., Huang, J., et al.: Liquid level intelligent detection of high voltage cable porcelain termination using short-time Fourier transform and deep brief networks. *China Measurement Testing Technol.* **45**(04), 47–52 (2019)
6. Liu, S., Xia, D., Qiu, Q., et al.: Researches on DC instantaneous over current impulse characteristics of flux coupling type pancake coil. *Adv. Technol. Electr. Eng. Energy* **38**(01), 47–53 (2019)
7. Yin, Z., Ning, Z.: Study on the test method of nondestructive stress test of space truss structure by magnetic flux method. *J. Disaster Prevention Mitigation Eng.* **40**(05), 803–810 (2020)
8. Liu, S., Li, Z., Zhang, Y., Cheng, X.: Introduction of key problems in long-distance learning and training. *Mob. Networks Appl.* **24**(1), 1–4 (2019). <https://doi.org/10.1007/s11036-018-1136-6>
9. Qi, Z., Li, X., Li, P., et al.: Fault detection method for cable joint based on DKPCA. *J. Dalian Univ. Technol.* **60**(03), 300–305 (2020)
10. Shenghui, L., Xue, B., Henan, D., et al.: Cable incipient fault identification based on stationary wavelet transform and random forest. *Adv. Technol. Electr. Eng. Energy* **39**(03), 40–48 (2020)
11. Liu, Z., Su, F., Wang, X., et al.: Partial discharge monitoring system based on cable terminal with distributed wireless TEV sensors. *Electr. Measur. Instrum.* **56**(17), 102–108 (2019)
12. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>
13. Zhou, L., Cao, J., Wang, S., et al.: Comprehensive state evaluation of high voltage cable based on multi-state variables characteristics and variation law. *High Voltage Eng.* **45**(12), 3954–3963 (2019)



# Design of Construction Cost Over Budget Control System Based on BIM Technology

Yun-xia Xie<sup>(✉)</sup> and Miao Wei

Guangxi Modern Polytechnic College, Hechi 547000, China  
rengjing65474@163.com

**Abstract.** The traditional cost over budget control system gets higher construction cost, and the deviation between budget value and actual value is large. In view of this problem, this paper designs a BIM technology-based construction project cost over budget control system. In terms of system hardware, the front-end acquisition device is optimized to measure the structural parameters of buildings; Optimize the real-time monitoring device of building construction and collect the construction images; Optimize the construction data integrator, integrate the measurement data and image data; In the aspect of system software, the lightweight database is configured according to the functional requirements of data integrator. After preprocessing the collected and monitored image data, the macro and micro influencing factors of construction cost are extracted. Finally, building construction model is established based on BIM Technology, and the influencing factors are displayed and adjusted to make the cost within the budget. The experimental results show that, compared with the traditional system, this system reduces the budget value of construction cost and the deviation between the budget value and the actual value.

**Keyword:** Project cost · BIM technology · Construction data · Budget value

## 1 Introduction

At present, there is a deep understanding of the cost management of construction projects in China, and relevant departments have been set up to carry out special project management audit. The intervention of government functional departments can more effectively complete the project management process. A professional cost management team is composed of more professional cost and accounting talents, which is bound by government functional departments, To supervise the process of project cost management, more effective project cost management, but due to the huge amount of project, the number of participating units and individuals is large, the project cost of construction project is still over budget. Therefore, the design of the construction project cost over budget control system and effective project cost control in the whole life cycle can not only save money for the country, but also achieve real-time supervision, timely discover the design risk and construction risk of the project, so that the project cost control runs through the whole life cycle of the project, which is of great significance to the design unit, the construction unit and the construction unit

construction units, supervision units and other departments of the project management ability also put forward higher requirements.

At present, the theory and practice of project cost management have reached the level of standardization, and a scientific and standardized management system has been established. Through practical cost methods, the benign process from quotation to implementation has been realized. The comprehensive cost control is defined as a set of theories and methods applied to cost management, and enterprises can use the methods of comprehensive cost management Program to manage the whole life cycle cost of all its investments, and manage all kinds of input resources of all strategic assets, including capital, time, human resources and material resources [1, 2]. In the whole life cycle cost management of enterprise strategic assets, we use the comprehensive method to carry out the whole process cost of all the resources invested, get the comprehensive cost management model, and realize the cost management of the whole life cycle of assets. The project life cycle includes four stages: project initiation, plan, implementation and completion.

On the basis of the above theory, this paper designs a BIM technology-based construction project cost over budget control system. An isolation link is added to the data integrator interface to transfer the associated integrated signal of the digital resource from the signal line in and out, and isolate the interference signal to ensure the stability of the communication signal during the associated integration of the digital resource. Configure a lightweight database according to the functional requirements of the data integrator, extract the macro- and micro-influencing factors of the construction project cost, establish a construction model based on BIM technology, display the influencing factors and adjust, so that the cost is within the budget.

## 2 System Design

### 2.1 System Hardware Design

#### **Optimize the Front-End Acquisition Device of Construction Data**

Through the integrated control equipment, the front-end acquisition device is optimized, and the building structural parameters are measured to provide data support for the control of construction cost exceeding budget. The overall structure of the system is composed of five parts: high voltage energy taking device, special frequency source, repeater, monitoring master station and front-end acquisition device.

The monitoring master station adopts C/S structure and is equipped with an IBM server to analyze and process the construction data. The client of the master station selects the PC and calls the IBM server data in real time to obtain the data of construction funds, time, human resources and material resources. Special frequency injection device is used to inject different frequency signal and communicate with master station directly.

The front-end acquisition device is installed in the construction position, powered by the repeater to detect the construction process and upload the detection data to the repeater. Through the high-voltage energy taking device, it supplies power to the

repeater, makes the repeater communicate with the master station, and sends the construction information data to the monitoring master station. The high voltage bushing is used in the structure of the energy taking device, and the device is built in the bushing. The high voltage metallized film capacitor is used to convert the high voltage into the low voltage, so as to realize the zero load start of the control system. The system communication consists of data transmission device and communication network. Two communication modes are selected for the communication network, namely GSM/GPRS communication and RF wireless networking. The frequency range of RF network is 350 kHz–28 GHz, and the transmission rates of GSM and GPRS are 1–65 kb/s and 1–120 KB/s respectively. The repeater uses MCU to control GSM module and GPRS module to communicate with front-end acquisition device and master station respectively. The master station uses GPRS module to receive data.

The integrated control equipment of front-end acquisition device integrates energy taking unit, switch, control unit, detection unit, measurement unit and circuit breaker. Using lightweight and miniaturized high-voltage capacitor, the power is directly taken from the wire as the working power supply of the integrated control equipment. The measurement unit is hung on the building, which is composed of voltage measurement unit, energy acquisition unit and single-chip microcomputer to collect construction data. The technical parameters of the front-end acquisition device are shown in Table 1.

**Table 1.** Technical parameters of front end acquisition device

Equipment	Parameter	Numerical value
Integrated control equipment	Rated input current	2 mA
	Load	<130 Ω
	Linear range	10–1800%ln
	Rated point ratio difference	<0.2%
	Rated output current	2 mA
	Working temperature	–40–90°C
	Accuracy class	0.1
	Permissible error	<±0.3%
Fault detector	Maximum rated voltage	11 kV
	Accuracy of load current measurement	<±1.5%
	Service voltage level	15–25 kV
	Applicable conductor diameter	10–50 mm
	working temperature	–20–80°C
	Altitude	<2000 m
	Communication distance	>150 m
	Rated frequency	60 Hz
Sampling resistance	120 Ω	

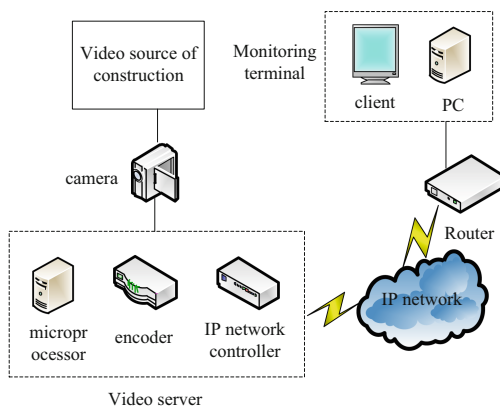
For the circuit breaker of the front-end acquisition device, the common vacuum circuit breaker, model ZW-72, is used. It is equipped with current transformer, isolation

knife, solar power supply, feeder automation terminal, intelligent data interface and combined transformer. Due to the large amount of installation work, the structure of the circuit breaker body is modified, the base is extended, the signal processing unit is installed on the base, the energy taking device is installed on the top of the extension section, and the signal processing unit is connected with the front-end acquisition device through the cable. The cable adopts twisted pair cable and is erected along the construction project. This method can ensure that the acquisition device has good anti electromagnetic ability, and ensure that the data acquisition and transmission are not interfered.

The front-end acquisition device adopts multi-channel interface, which enables automatic switching between acquisition channels, including digital signal, analog signal, network signal, GPRS signal, etc., monitors each channel, analyzes the detection data of integrated control equipment and fault judgment device, and converts it into the actual value of construction state [3]. So far, the optimization of the front-end acquisition device of construction data has been completed.

**Optimization of Real Time Monitoring Device for Building Construction**

Optimize the construction real-time monitoring device, carry out real-time monitoring on the construction project, and collect the construction image. The real-time monitoring device for building construction is composed of three parts: video acquisition device, video server and monitoring terminal. The USB camera is used as the video acquisition device, and the model au4660 is selected. At the front end of the camera, the image clarity ball device is configured to ensure the clarity of the collected image. The hardware of video server is based on SBC2440 embedded platform, embedded with Samsung S3C2440 microprocessor, connected with acquisition equipment and server through USB1.1 bus, and installed in the monitoring site of construction engineering [4]. The monitoring terminal adopts PC, and the overall architecture of the monitoring terminal is shown in Fig. 1.



**Fig. 1.** Overall structure of monitoring terminal

USB camera is used to read the video image information of the construction project, compound the analog video signal, and transmit the digital signal of the video to the video server after a/D conversion. The analog-to-digital converter is set with a precision of 12 bits and a speed of 25 MSPs. The internal analog video switch is configured. The conversion circuit between analog video signal and digital video signal is shown in Fig. 2.

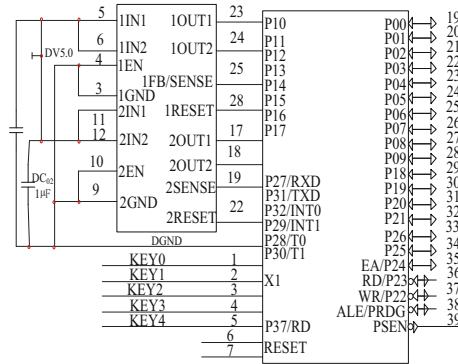


Fig. 2. Conversion circuit of video analog signal and digital signal

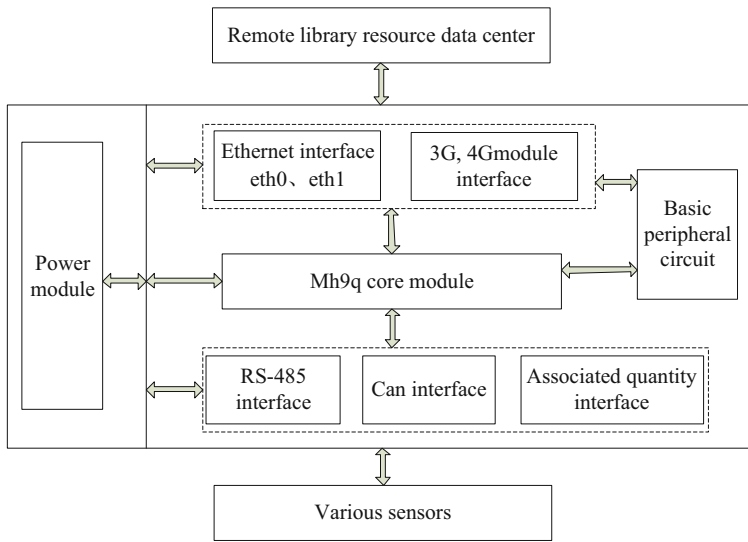
In Fig. 2, the left side is the analog-to-digital conversion circuit, and the right side is the composite video processing circuit. Through the composite processing circuit, the chroma and brightness of the video signal are separated, and the parallel YCbCr signal is output. The development board of video server integrates serial interface, SD card storage interface, USB interface, 1.5 mm hard disk interface and IP network interface. The hard disk interface meets the external standard of microprocessor and encoder. S9c1860 is selected as the server chip. The core of the chip is a high-speed buffer structure with tailorability. It can make full use of the hardware resources on the server, and the maximum frequency can reach 540 MHz, which can meet the application requirements of Embedded Server [5]. Run AUS729 encoder on the server to compress the video image to form H.264 video stream in YCbCr format. Through the network layer interface of IP network controller, encapsulate the compressed video stream, convert it into RTP format packet, and send the video stream to the monitoring end through IP network multicast. The monitoring terminal decodes the video stream according to the set format, decodes the H.264 stream into the original image in YCbCr format, and displays the monitoring video of building construction. So far, the optimization of real-time monitoring device for building construction has been completed.

**Optimizing Construction Data Integrator**

Optimize the construction data integrator, integrate the measurement data and image data. According to the service function of the construction project cost over budget control system, the data integrator of the system is optimized, and the overall construction of the integrator is divided into five modules: core, associated interface, power supply, network function and peripheral circuit. SVB12F762 chip is selected for the



integrator, and its configuration should meet the circuit requirements of the integrator. SVB12F762 chip is used for level conversion, so that the integrator bus can maintain high level and weaken the reflected signal when it is idle. The isolation link is added in the data integrator interface to transmit the correlation integration signal of digital resources from the signal line, and isolate the interference signal, so as to ensure the stability of the communication signal in the process of correlation integration of digital resources [6]. The overall framework of its data integrator is shown in Fig. 3.



**Fig. 3.** Framework of data integrator

As shown in Fig. 3, the core module adopts MH9Q core platform produced by Freescale company to build 4 GB eMMC Flash and 2 GB memory to integrate the construction data, integrate it into Cortex-a9 four core processor, and drive the resource data through Ethernet and CAN bus, so as to support the operation of data association and integration. In the optimization of the power module, the data integrator uses 24 VDC power input to make the collector face the industrial general interface. The power module is connected with SVB12F762 chip to convert the transistor logic level signal. The level is directly connected with the input and output interfaces of the core module, and the CAN transceiver of the construction data is added to make it output 5 V power supply with stable voltage, so as to realize the overall optimization of the data integrator. So far, the optimization of construction data integrator is completed, and the hardware design of the system is realized.

## 2.2 System Software Design

### Design System Database

According to the functional requirements of data integrator, a lightweight database is configured to store the collected data orderly and record the operation status of construction. DAQ is embedded in the hardware of the `daq_status_tb` table and `data_store_tb` table, where `daq_status_tb` is used to store the running state of construction, `data_store_tb` is used to store the data of building structure parameters. The specific system database configuration is shown in Table 2.

**Table 2.** System database configuration

Field name	Field meaning	Field type
<code>start_time</code>	Serial number of data association integration interface	Blob
<code>sensor_num</code>	System operation status	Interger
ID	Ms level, time stamp	Interger
<code>daq_serial</code>	Autoincrement primary key	Interger
Status	Serial number of various sensors	Text
<code>Raw_data</code>	Construction data association and integration serial number	Text
<code>sensor_number</code>	raw data	Interger
<code>store_date</code>	Association type	Blob
<code>daq_type</code>	Integrated data	Text
<code>Process_data</code>	Autoincrement primary key	Interger

Using Mysql database and concurrent operation, we can directly access the data files of construction. In the PHP processing module of the database, the user uses `text / html; Charset = UTF-8 //` sends the access request, then processes the php script, parses the data of book digital resources, reads and writes the data configuration file through `login-btn > submit-text2//`, receives the information returned by the php script, and finally generates `var-export (list1,10) -catch//` to present the data to the whole web page, so as to display the data in the database.

At the same time, Ontology semantic web technology is used to realize the digitization of the storage resources in the database, describe the construction data, get the descriptive metadata, and send it to the network as the associated data. According to the code of Cool URIs named by semantic web, the URI of construction data is named. With the help of various description methods provided by FRBR, the associated data vocabulary set is created to describe the semantic ontology of digital resources. The semantic classification standard of Digital Resource Association is shown in Table 3.

**Table 3.** Classification standard of digital resource association semantics

Semantic name	Subnet type involved	Segmentation criteria
Hierarchy	P-P;K-K;M-M	Genus
Citation relation	P-P;M-M	Entity
Correlation	P-P;K-K;M-M	Whole part
Equivalence relation	P-P;K-K;M-M	synonymous
Attribute relation	P-P;P-K;K-K;M-M	Near meaning
Discuss the relationship	K-K;K-M;M-M	synonym
See relation	P-K;K-K;K-M;M-M	Antisense

According to the content shown in Table 3, the progressive transformation mechanism of construction data association semantics is constructed, and the corresponding subnet type is selected. Then, the associated semantics of construction data is described. With the help of entity extraction mechanism and D2RQ conversion tool, the construction data is transformed into RDF metadata. On this basis, a new semantic descriptive metadata is created [7]. Finally, the publication mode of associated data is selected to expand the construction data, collect the Open API provided by construction engineering, build a network environment with stronger resource correlation and scheduling, and combine services and associated data to connect the control system with internal and external construction data. So far, the design of the system database is completed.

### Preprocessing Construction Data

Preprocess the image data collected by acquisition device and monitoring device, remove data noise and reduce memory space. Through weighted mean filtering, the noise of video image is removed, the weight of each pixel is set, and it is judged that the closer the pixel is to the center point, the larger the weight is, and the farther the pixel is to the center point, the smaller the weight is, so as to reduce the blurring of video image [8–10]. Select a fixed size local window, calculate the gray value of each pixel in the window and the average gray value of neighboring pixels, and replace the gray value with the average gray value. The expression of weighted mean filter is as follows

$$f(x, y) = \frac{\sum_{i=0}^n R_i(x, y)w_i(x, y)}{\sum_{i=0}^n R_i(x, y)} \quad (1)$$

where,  $(x, y)$  is the pixel coordinates of the video image,  $n$  is the number of center points of the video image,  $w_i(x, y)$  is the gray value of the pixel in the neighborhood of the center point of the filter window,  $R_i(x, y)$  is the corresponding weight of  $w_i(x, y)$ , and  $f(x, y)$  is the gray value of the center pixel after filtering. The formula (1) is used to improve the digital signal of video image edge and suppress the influence of noise

pixels on neighboring pixels. According to the gray value, The  $h$  expression of gray image channel is as follows

$$h = (0.114 \quad 0.587 \quad 0.299) \begin{pmatrix} B \\ G \\ U \end{pmatrix} \quad (2)$$

Among them,  $B$ ,  $G$  and  $U$  are blue, green and red channels respectively, and 0.114, 0.587 and 0.299 are channel threshold parameters respectively. The expression formula of threshold segmentation is as follows

$$J(x, y) = M(h, Q) \quad (3)$$

Among them,  $M$  is Gaussian matrix,  $Q$  is channel threshold index of video image, and  $J(x, y)$  is the threshold segmentation video image. Through formula (3), the multi-pixel width of the image line is reduced to the unit pixel width, and the structural information of the features is stored in the image to realize the thinning of the video image, so as to recognize the features contained in the image, extract the feature range of the video image, reduce the data structure of the H.264 video stream, and reduce the image data operation efficiency. So far, the pretreatment of construction data is completed.

### Mining Video Frame of Construction Project

Using big data tag technology, mining construction video frames in the preprocessed image data. According to the extracted image contrast, structure information and brightness, the video image is compared, and the similarity of the preprocessed image is calculated:

$$\begin{cases} j(a, b) = (2u_a u_b) / (u_a^2 + u_b^2) \\ p(a, b) = (2s_a s_b) / (s_a^2 + s_b^2) \\ g(a, b) = (2s_{ab}) / (s_a s_b) \end{cases} \quad (4)$$

Among them,  $a$  is the preprocessed video image,  $b$  is the contrast video frame,  $j(a, b)$ ,  $p(a, b)$  and  $g(a, b)$  are the brightness similarity, contrast similarity and structure similarity of the video frame,  $u_a$  and  $u_b$  are the average brightness values of  $a$  and  $b$  frames,  $s_a$  and  $s_b$  are the standard deviation of  $a$  and  $b$  frames, and  $s_{ab}$  is the covariance of two frames. Then the final similarity  $Z$  of the video frame is:

$$Z = [j(a, b)]^m [p(a, b)]^n [g(a, b)]^o \quad (5)$$

Among them,  $m$ ,  $n$  and  $o$  are the importance proportion of the three modules. In the H.264 video stream, the video frame with high final similarity is selected as the associated data in the massive video data. Using the big data label technology, the data label is established, and the associated image data is clustered to label each cluster of data. According to the tags of all the data, it is classified and stored in the tag library,

and the source of video frames is analyzed. The alarm information of tag library is shown in Table 4.

**Table 4.** Label library alarm information table

Data type	Listing	Space	Data description
User number	varchar	yes	Construction time
User name	varchar	yes	Construction address
Logon times	varchar	no	Construction number
describe	char	no	describe
Last logon time	varchar		Linkage intercom
Alarm number	datetime	yes	construction stage
Alarm name	varchar	yes	Construction structure
Alarm level	varchar	yes	construction technology
Alarm status	datetime	yes	Construction time
Occurrence time	varchar	no	Construction funds
End time	datetime	yes	Construction manpower
Alarm event	datetime	no	Construction material resources
Linkage intercom	char	no	Construction status
Remarks	varchar	no	Completion time of construction

For the newly added construction video images, the video frames are allocated to the most similar clusters, which are also transmitted to the tag library to search for data tags. In the massive monitoring video image data, the construction video frames are mined out, and after exceeding the budget control, the video frames are labeled with invalid Tags. So far, the excavation of construction video frame is completed.

### **Extraction of Macro Influencing Factors of Construction Cost**

In the excavation of construction video frame, the influencing factors of project cost exceeding budget are extracted. The Macro Influencing Factors of cost are summarized as follows:

1. The total length of the building, when the length of the building is long, the expansion joint needs to be set. The length of the frame structure is more than 55mm, and the length of the shear wall structure is more than 45mm. The shrinkage stress and temperature stress inside the concrete of the super long structure are relatively large, so the amount of reinforcement will naturally increase, and it is also unfavorable to the earthquake resistance of the structure, In the case of no expansion joint, expansion post cast strip is usually set with a width of 2 mm, and additional reinforcement is set in the strip, which will also increase the amount of steel used;
2. For buildings with large plane aspect ratio and large plane aspect ratio, due to the large difference of dynamic characteristics in the two principal axis directions, under the horizontal earthquake, the stress of components in the two directions is not uniform, and the structural torsion coefficient is large, which leads to the increase of component reinforcement;

3. Vertical aspect ratio, for high-rise buildings, the overturning moment of the structure with large aspect ratio is larger, which leads to the poor overall stability of the structure. In order to ensure the overall stability of the structure and control the lateral displacement of the structure, it is necessary to increase the lateral force resisting members or improve the stiffness of the lateral force resisting members, which will increase the amount of concrete and reinforcement, thus increasing the cost;
4. The buildings with regular and uniform facade shape and vertical shape have no sudden change in stiffness, even stress and small steel consumption. If there is vertical irregularity, such as cantilever structure, the calculation reinforcement will be increased in stress, and the requirements for seismic structure will be higher, so as to increase the cost;
5. Plane shape: the simpler the plane shape of a building is, the better the stress is, the less the steel consumption is. Irregular plane shape not only increases the steel consumption, but also has a great impact on the seismic performance of the structure;
6. The reasonable layout of column net and the size of column net can reduce the stress of components and the steel consumption;
7. Different types of buildings, such as villas, multi-storey buildings, small high-rise buildings, high-rise buildings and super high-rise buildings, all have different design and construction specifications. For example, the design specifications of small high-rise buildings and super high-rise buildings are different, which are reflected in the aspects of earthquake resistance, wind resistance, fire prevention and pressure increase;
8. Floor height, on the premise of meeting the building function, appropriately reducing the floor height will reduce the project cost. For every 10 cm decrease in floor height, the project cost will be reduced by about 1%, and the wall materials will be saved by about 10%;
9. Lateral force resisting system, because the horizontal force in high-rise structure has a great influence on the structure, so lateral force resisting system becomes an important part of high-rise structure system. The amount of structural materials used to bear vertical load in unit building area is linearly proportional to the number of floors of the building. When the floor load is determined, the reinforcement of the floor structure is certain, which has almost no relationship with the number of floors of the structure. The material consumption of vertical load-bearing components such as walls and columns increases linearly with the number of floors of the building. The structural material consumption for resisting lateral force increases in a quadratic curve according to the number of floors of the building. Is the selection of lateral force resisting system economical and reasonable, Directly affect the amount of structural materials [11].

So far, the extraction of macro factors affecting the construction cost has been completed.

### **Extraction of Micro Influencing Factors of Construction Cost**

In the mining video frame, the micro influencing factors of project cost exceeding budget are extracted as follows:

1. Generally speaking, for the residence with more than 8 floors, the shear wall structure adopts more structural forms. When the shear wall layout is reasonable and the limb length is appropriate, most of the reinforcement of the section can be arranged according to the structure. According to the technical specification for concrete structure of high rise buildings, the seismic grade of short limb shear wall with the wall length of 5–8 times, Therefore, in the plane layout, the number of short pier walls should be reduced as much as possible, and the amount of structural reinforcement should be reduced accordingly;
2. In the plane layout of shear wall structure, the windows at the corner should be reduced as much as possible, and the shear wall at the corner should be retained to control the torsion of the whole building. The more the wall limb at the corner is weakened, the more the amount of concrete needed for other components is, and the corresponding reinforcement of beams and walls will also be increased;
3. The torsion effect of the structure can be reduced if the position of the lateral force resisting member and the center of stiffness coincide with or close to the center of mass. By reasonably arranging the lateral force resisting member, the torsion can be effectively adjusted, so as to reduce the steel consumption;
4. In the layout of beams, slabs, columns, walls and other components, the reasonable stress and direct force transmission should be considered to make the transfer path of floor load as simple as possible, avoid the unclear force transmission of multi-level beams, and then meet the requirements of normal use. For the case of partition wall on the slab, the load of partition wall can be changed into uniform load on the floor, and the method of adding reinforcement at the bottom of the slab is adopted instead of setting beams to increase the cost;
5. The section size of the wall and column should be reasonably determined in the structural design. The wall and column are compression bending members, and the actual reinforcement amount is generally in accordance with the structural requirements. On the premise of meeting the specification, the axial compression ratio of the column should not be too small, and the section size of the column is too large, which not only affects the use of indoor space, but also increases the foundation size and cost;
6. The reinforced part of the shear wall, the first and second shear walls are divided into the bottom reinforced part and the non reinforced part. The reinforced part is reinforced according to the constraint edge member, and the non reinforced part is reinforced according to the structural edge member. The reinforcement amount of the nodes in the bottom reinforced area and the edge member is much larger than that of the non reinforced part, so the structural design should be in accordance with the specification, Strictly distinguish the seismic grade and shear wall position, so as to avoid unnecessary waste;
7. On the premise of meeting the structural design, the selection of high-grade reinforcement scheme can not only reduce the project cost, but also increase the building safety reserve and concrete structure strength, especially for high-rise and important buildings;

8. On the premise of meeting the requirements of the code, reducing the weight of the structure, reducing the number of shear walls, controlling the stiffness of the structure, thinning the shear wall, changing the windowsill wall into brick, opening the long wall, and designing the large bay can achieve good benefits.
9. The choice of foundation form is not only related to the direct engineering cost of buildings, but also affects the foundation pit maintenance.

So far, the extraction of micro influencing factors of construction cost is completed.

### **Control of Construction Cost Exceeding Budget Based on BIM Technology**

The BIM Technology is used to establish the building model, and the macro and micro factors are displayed through the model to control the over budget of the construction cost. In the process of construction, CAD drawings are imported into 3D-BIM modeling software to create real component information to avoid drawing recognition errors due to the long time of drawing review. Then, a visual model is generated by adding the corresponding coordinate information artificially. When managing the construction progress, the collected construction data are added to obtain the four-dimensional model of the construction project. For the civil engineering and installation part, the technical department uses the MagiCAD software to complete the modeling of the mechanical and electrical part, and the software engaged in the calculation work is used for the cost budget part. Through the construction network provided by BIM software, it helps CAD drawings express the building space intuitively, making the construction progress more visible and easy to track the progress of the Project. Meanwhile, the project schedule prepared by project software reflects the space and time information in the construction, and facilitates the control of the construction period [12, 13].

This study applies BIM Technology to the field management. Through the model and field simulation, we can know the possible events in the construction site in advance. The accurate database provided by BIM Technology model created by the project is used to track the site, so as to achieve the effect of dynamic management of the factors affecting the project cost.

At this point, the cost of construction project based on BIM Technology is over budget control, and the software design of the system is realized. Combined with hardware design and software design, the design of the construction cost over budget control system based on BIM Technology is completed.

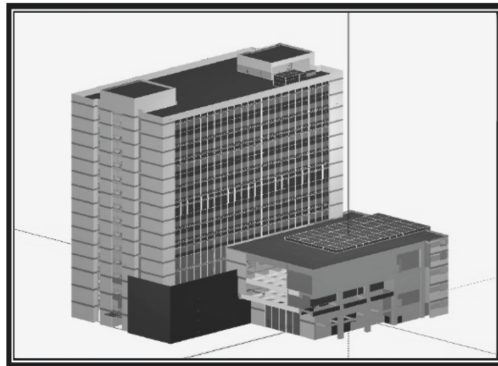
## **3 Experiment and Analysis**

In order to verify the reliability of the BIM technology-based construction project cost over budget control system, the traditional construction project cost over budget control system is selected as the contrast, and the budget value and the deviation between the budget value and the actual value of the construction project under the control of the two groups of systems are compared.



### 3.1 Summary of Construction Engineering

The construction project is an office building project with a building height of 22.5 m and a building area of 9210.47 m<sup>2</sup>. The building base area is 1490.57 M<sup>2</sup>. The building has two floors underground and 13 floors above ground. The second floor is the underground garage, and the first floor is the basement. Office and technical buildings are located on the 13th floor above ground. The second floor underground is a fire compartment. The first floor underground has two fire compartments. The first floor and the second floor are one fire compartment. One fire compartment is set on the third floor to the thirteenth floor. The structural type of the building is frame structure, the seismic grade is grade 3, the seismic structural measures are grade 2, and the structural safety grade is grade 2. The seismic fortification category is class B, the design grade of foundation is class C, the environmental category of concrete structure is class 2b, the environmental category of indoor concrete structure is class I, and the environmental category of toilet is class 2a. The BIM model of the design system is shown in Fig. 4.



**Fig. 4.** Building BIM model

### 3.2 Experimental Result

The two groups of systems are used to control the construction project cost exceeding the budget, and the budget value of the construction project cost under the control of the two groups of systems is compared. The results are shown in Table 5.

**Table 5.** Comparison results of construction project cost budget (10000 yuan)

	Construction stage	This paper system	Traditional system
Underground	1st floor	8392.40	8976.20
	2nd floor	8251.50	8816.40
On the ground	1st floor	7936.20	8826.10
	2nd floor	8142.40	8991.90
	3 layers	8028.30	8718.60
	4th floor	7826.10	8834.50
	5 floors	7935.90	8971.40
	6th floor	8092.70	8816.30
	7th floor	8271.20	8782.90
	8th floor	8172.60	8826.20
	9th floor	8204.40	8803.40
	10th floor	7906.90	8714.50
	11th floor	7836.40	8829.10
12th floor	8152.60	8902.30	
13th floor	8071.40	8816.40	

It can be seen from Table 5 that the construction cost of the system budget in this paper is significantly lower than that of the traditional system. In this paper, the total budget value of the system construction project cost is 1212.21 million yuan, and the total budget value of the traditional system construction project cost is 1326.262 million yuan. Compared with the traditional system, the total budget value of the construction project cost of this system is reduced by 114.052 million yuan, which proves that it saves more construction cost.

**Table 6.** Comparison results of project cost budget deviation (%)

	Construction stage	This paper system	Traditional system
Underground	1st floor	0.09	0.26
	2nd floor	0.08	0.29
On the ground	1st floor	0.08	0.25
	2nd floor	0.07	0.19
	3 layers	0.06	0.25
	4th floor	0.09	0.28
	5 floors	0.05	0.22
	6th floor	0.08	0.25
	7th floor	0.06	0.24
	8th floor	0.08	0.12
	9th floor	0.12	0.29
	10th floor	0.11	0.21
	11th floor	0.06	0.18
12th floor	0.07	0.22	
13th floor	0.10	0.23	

On this basis, the deviation between the budget value and the actual value of the project cost obtained by the two systems is compared. The experimental results are shown in Table 6.

It can be seen from Table 6 that the construction cost obtained from the system budget in this paper is closer to the actual value. The average deviation of the construction project cost of the system budget in this paper is 0.08%, and that of the traditional system budget is 0.23%. Compared with the traditional system, this system budget construction cost deviation, reduced by 0.15%, the budget value is more accurate and reliable. The main reason is that the method in this paper adds an isolation link to the data integrator interface, transmits the associated integrated signal of the digital resource from the signal line, and isolates the interference signal to ensure the stability of the communication signal during the associated integration of the digital resource.

## 4 Conclusion

This study gives full play to the modeling advantages of BIM Technology, reduces the budget value of construction cost, and reduces the deviation between the budget value and the actual value. However, there are still some deficiencies in this study. It takes a lot of time to establish BIM model for engineering quantity calculation. In the future research, it will simplify the calculation process of reinforcement and civil engineering, and further improve the control accuracy of construction cost over budget.

### Fund Projects

1. Guangxi Vocational Education Teaching Reform Research Project in 2020 “Research and Practice on the Practice Teaching System of “Three Courses and Four Abilities” in Higher Vocational Architecture Majors with Progressive Ability” (GXGZJG2020B122)
2. In 2020, Guangxi Gao School's young and middle-aged teachers' scientific research ability improvement project “Research and Practice of Prefabricated Buildings Based on BIM Technology in Underdeveloped Areas” (2020KY45013).

## References

1. Lei, C.: Analysis and control measures of overbudgeting of construction cost. *Shanxi Architecture* **45**(3), 221–222 (2019)
2. He, S., Zhuo, L.: Study on the causes and control of landscape engineering costs beyond the budget. *Urbanism Architecture* **16**(15), 185–186 (2019)
3. Yu, N.: Reasons for over budget of construction project cost and control countermeasures. *Build. Techn. Dev.* **46**(7), 100–101 (2019)
4. Chen, L., Liang, K.: Cause Analysis and control method exploration of project settlement cost exceeding bbudget. *Equipment Manufacturing Technol.* (10), 113–114+149 (2019)
5. Yuan, Y.: Analysis of the causes and control measures of construction project over budget. *Value Eng.* **39**(6), 16–17 (2020)

6. Cao, W.: Causes and control methods of construction cost over budget. *Const. Design Project* (13), 252–254 (2020)
7. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective Template Update Mechanism in Visual Tracking with Background Clutter. *Neurocomputing* (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>
8. Li, J.: Analysis of the causes and control strategies of overbudget of construction cost. *Const. Design Project* (12), 245–246 (2019)
9. Liu, S., Li, Z., Zhang, Y., Cheng, X.: Introduction of key problems in long-distance learning and training. *Mob. Networks Appl.* **24**(1), 1–4 (2018). <https://doi.org/10.1007/s11036-018-1136-6>
10. Ma, X., Liu, D.: 3D reconstruction simulation of BIM building based on perspective augmented reality. *Comput. Simul.* **37**(3), 229–233 (2020)
11. Wang, D.: Analysis of the causes and control measures of construction project cost exceeding budget. *Manage. Technol. SME* (8), 86–88 (2020)
12. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>
13. Yue, L.: Application of BIM technology in installation budget. *Eng. Cost Manage.* (5), 50–56 (2019)



# Cluster Evaluation Method of Building Decoration Cost Rationality Based on BIM Model

Miao Wei<sup>(✉)</sup> and Yun-xia Xie

Guangxi Modern Polytechnic College, Hechi 547000, China  
wow8858@aliyun.com

**Abstract.** In view of the low efficiency of traditional cost rationality evaluation methods, this paper designs a cluster evaluation method of building decoration cost rationality based on BIM model. BIM model is used to collect the data of architectural decoration, and the reasonable threshold interval is set. Then the decoration data and threshold parameters are clustered to complete the construction of the evaluation method. After setting the classical domain of quantitative indicators, we prepare the evaluation data set for the experiment, and call two traditional evaluation methods and the evaluation method designed in this paper to carry out the experiment. The results show that the evaluation method designed in this paper is more efficient.

**Keyword:** BIM model · Building decoration cost · Rational clustering · Evaluation method

## 1 Introduction

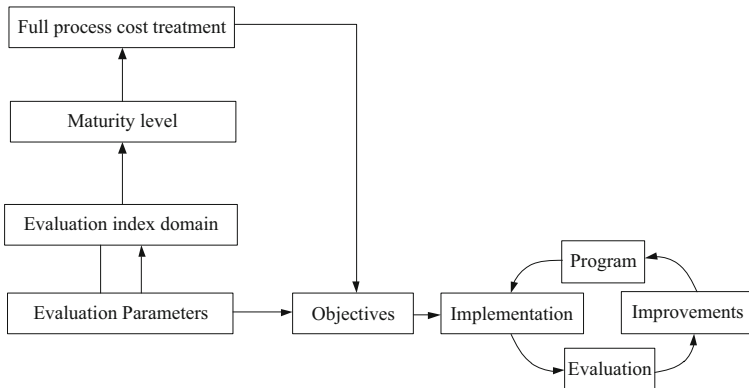
The rapid growth of the market is bound to attract fierce competition from enterprises, the core of the competition is how to maximize the profit, how to maximize the profit is the current stage of the construction decoration enterprises facing the need to urgently solve the key problem. The background in the same environment, in addition to the combination of market demand changes to adapt to the product development, in the process of implementation of project construction, which enterprises can guarantee under the premise of construction schedule, quality, cost of the project's development have enough attention, to control the cost in a reasonable manner, with the lowest cost, realize own benefit and social benefit as much as possible, Then the enterprise can grasp the initiative of market competition more, and realize the survival and development [1–3]. The implementation of architectural decoration has a complete periodicity. First of all, in the investment stage, how to grasp the accuracy of investment directly determines the profit and loss of the later project. In the design stage, controlling the design standard can avoid too many changes and claims in the later stage. In the bidding stage, the false quotation of the contractor will also lead to the difficulty of real estate cost control [4]. In the process of construction, how to combine with the actual situation of construction, reasonable preparation of cost scheme, will also affect the final profit to a greater extent.

Therefore, in view of the above background, this paper designs a cluster evaluation method of building decoration cost rationality based on BIM model. This paper uses the BIM model to collect architectural decoration data and sets the reasonableness threshold interval. Then the decoration data and threshold parameters are clustered to complete the construction of the evaluation method.

## 2 Method Design

### 2.1 Using BIM Model to Collect Building Decoration Data

When using BIM model to collect architectural decoration data, during the whole construction cycle of the project, different stages of work are completed by different participants, and different objectives and interests of each participant will produce different types of cost management methods. At this time, the owner who is in the leading position of project management is particularly important to the management of the overall project cost. The owner's project cost management runs through the whole process of the project, including investment decision-making stage, design stage, bidding stage, implementation stage and completion stage [5, 6]. Under the control of BIM model, the data collection model formed is shown in Fig. 1.



**Fig. 1.** Data collection process of BIM model

In the data collection process shown in the figure above, the maturity level corresponding to each stage of the whole process of project cost management is gradually decomposed into evaluation index field, which contains several evaluation parameters that can be operated in practice. The basic logic of project cost management maturity model is to calculate the maturity level of each evaluation parameter:

$$g = p \frac{Z}{\alpha} \tag{1}$$

Among them,  $g$  is maturity level,  $p$  is evaluation parameter,  $Z$  is numerical parameter and  $\alpha$  is basic parameter [7, 8]. Under the control of the above numerical relationship, the maturity level of project cost management is comprehensively calculated. By setting goals for the evaluation parameters of each stage, the cost management is refined into the basic activities of each process. The maturity model of project cost management specifies the objectives for each evaluation parameter, and the objectives summarize the content of basic activities, and can be used as the judgment standard of whether an organization or project has effectively implemented basic activities. By measuring the evaluation parameters of each stage, the maturity level is obtained [9, 10].

Based on the on-site investigation of architectural decoration cost staff and consulting units, and interviews with industry experts and experienced management personnel, the initial evaluation index system is sorted out by analyzing the logical relationship between each index based on the extensive collection of key influencing factors. The index system is shown in Table 1.

**Table 1.** Index system

Decorative criteria layer	Index name	Evaluation parameters	
Investment decision stage	Project decision preparation	Adequacy of decision making preparation	
	Investment estimation	Preparation of investment estimation documents	
		Deviation rate of investment estimation	
	Risk management	Risk identification	
		Risk evaluation	
		Risk prevention measures	
	Investment and financing	Capital ratio	
		Financing cost ratio	
	Design phase	Preliminary design	Preparation of design budget estimate documents
			Deviation rate of Budget Estimate Compilation
Construction drawing design		Preparation of construction drawing budget document	
		Deviation rate of budget compilation and review	
		Quota Design	
Design Management		Standard design degree	
		Comparison and selection of design schemes	

*(continued)*

**Table 1.** (continued)

Decorative criteria layer	Index name	Evaluation parameters
Bidding stage	Bidding management	Standardization of bidding management
	Tender documents	Preparation of bidding documents
		Preparation of bill of quantities
	Preparation of the base price	Rationality of base bid preparation
Contract signing		Contract preparation
		Deviation rate of contract price
Implementation phase	Engineering change management	Adequacy of change basis
		Changes caused by bidding documents
	Claim management	Claim cost ratio
		Claim compliance
	Contract management	Contract management compliance
		Validity of performance of contract documents
	Project settlement	Preparation of settlement documents
		Settlement price deviation rate
		Preparation of final account documents for completion
	Project Acceptance	Final account of completion

Corresponding to the index system constructed in Table 1, the collected data are corresponding to different building decoration stages. After collecting cost data, the reasonable threshold range is set.

**2.2 Set Reasonable Threshold Range**

After using BIM model to collect data, when setting reasonable threshold interval, the method divides the time series into equal length segments, and uses the average value of data points in the segment to represent the time series. The longer each segment is divided, the larger the dimension reduction is, but the more information is lost. If each segment is too short, the effect of dimension reduction will not be obvious. In order to further retain the key information of time series, it is sorted into a standard series. After processing, dimension reduction is used to process the information into a piecewise function:

$$\bar{x} = \frac{w}{n} \sum_{j=n+1}^w x_j \tag{2}$$

where  $w$  is the piecewise value,  $n$  is the dimension reduction parameter, and  $x_j$  is the piecewise function [11–13]. According to the specified letter set size, the Gaussian



distribution table is used to find the interval split point. The time series is divided into equal length segments, and the middle point of each segment is selected. Then, the selected data points will be converted into a binary sequence, where 1 represents above the average value of the data points and 0 represents below the average value of the data points. The numerical distribution is shown in Fig. 2.

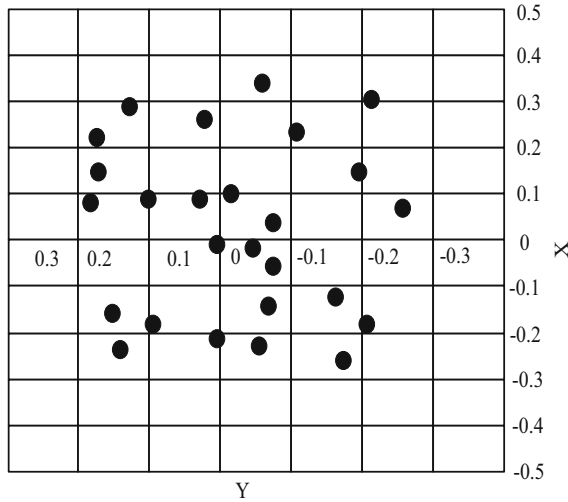


Fig. 2. Numerical distribution

Under the numerical distribution shown in the figure above, and the values in the time series are converted into binary series, it can be expressed as:

$$b_i = \begin{cases} 1 & \text{if } t \geq \mu \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where  $t$  is the conversion period and  $\mu$  is the average value of the time series. In the calibrated interval, the calculation formula of the numerical point distance in the conversion sequence can be expressed as:

$$D = \frac{\sqrt{\sum_{i=1}^N w_i + \sum_{j=1}^t d_j}}{b_i} \tag{4}$$

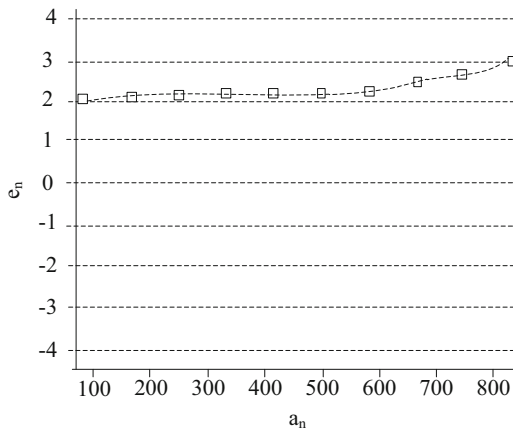
where  $w_i$  is the partition function,  $d_j$  is the numerical point function, and  $N$  and  $t$  are the sequence mean. In time series fitting, abnormal data will cause residuals do not conform to the normal distribution, so the analysis of residual series can determine the outlier model. The fitting residual is defined as:

$$\pi(B) = \frac{\theta(B)}{\varphi(B)} \tag{5}$$

where,  $\theta(B)$  is the normal residual value,  $\varphi(B)$  is the abnormal residual value, and the meaning of other parameters remains unchanged. In order to eliminate the abnormal data in the fitting residuals, the matrix method is used to expand the residuals:

$$\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \omega \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \tag{6}$$

Among them,  $a_n$  is white noise sequence,  $e_n$  is fitting value after expansion processing, and  $\omega$  is migration operator. In the process of fitting, the deviation is mainly caused by the outliers, and the values in the noise points are integrated into the test statistics of the noise points. The change of test statistics is shown in Fig. 3.



**Fig. 3.** Test statistics

Under the change of test statistics shown in the figure above, if the test statistics at any time exceeds the limit value, it is judged that there are noise points at that time. The value is taken as the final rationality threshold, and the limit value is taken as the control condition. Clustering process the relationship between decoration data and threshold parameters, and finally build the evaluation method.

### 2.3 Clustering Processing Decoration Data and Threshold Parameters

In the reasonable threshold interval, the cluster vectors in the interval are calibrated, and the labeling process can be expressed as follows:

$$AC = \frac{1}{N} \sum_{j=1}^t \max_k |\omega_k| \quad (7)$$

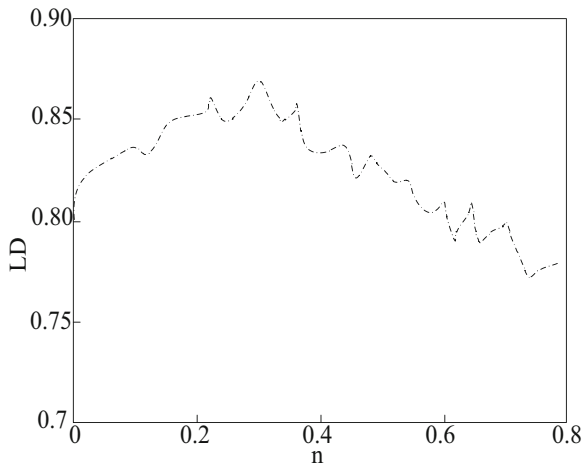
Among them,  $k$  is the cluster vector parameter,  $AC$  is the tag function, and the meaning of other parameters remains unchanged. After label processing, corresponding to different mutual information, intersection processing can be expressed as mutual function:

$$H(c) = \frac{|\omega_k \cap c_j|}{N} \quad (8)$$

where  $H(c)$  is the mutual function and  $c_j$  is the cluster function. In this information measurement process, the correlation between decoration data and threshold parameters is defined, and the RAND index formed by numerical clustering is calculated:

$$LD = \frac{2(a+b)}{n(n-1)} \quad (9)$$

Among them,  $a$  and  $b$  represent information entropy, and the meaning of other parameters remains unchanged. At this time, the index changes as shown in the figure below.



**Fig. 4.** Change of index value

Under the change of the value shown in the figure above. The introduction of an automatic encoder, the automatic encoder has a strong nonlinear expression ability, which may also fully describe the unique characteristics of individual samples, which results in the over fitting of input data. However, the structure of high-dimensional data

is different, and there are many special features. If the automatic encoder is directly used in high-dimensional data for dimensionalization. The possible low dimension eigenvectors can not reflect the public distribution characteristics of high-dimensional data, which leads to the poor generalization ability of the model obtained by training, and it is impossible to popularize it on other data. In order to reduce the weight amplitude and prevent over fitting, L2 norm regular term is added. The numerical relation can be expressed as follows:

$$J(w, b) = \frac{\lambda}{2} \sum_{j=1}^r n_j \tag{10}$$

where,  $n_j$  is the weight amplitude parameter, and the meaning of other parameters remains unchanged. Corresponding to the above numerical relationship, corresponding to the clustering parameters formed by different parameters, the parameters are shown in Table 2.

**Table 2.** Cluster parameters formed

Index name	Evaluation parameters	Cluster parameters
Project decision preparation	Adequacy of decision making preparation	0.316
Investment estimation	Preparation of investment estimation documents	0.266
	Deviation rate of investment estimation	0.338
Risk management	Risk identification	0.246
	Risk evaluation	0.258
	Risk prevention measures	0.205
Investment and financing	Capital ratio	0.282
	Financing cost ratio	0.239
Preliminary design	Preparation of design budget estimate documents	0.314
	Deviation rate of Budget Estimate Compilation	0.336
Construction drawing design	Preparation of construction drawing budget document	0.343
	Deviation rate of budget compilation and review	0.234
	Quota Design	0.282
Design Management	Standard design degree	0.334
	Comparison and selection of design schemes	0.334

(continued)

**Table 2.** (continued)

Index name	Evaluation parameters	Cluster parameters
Bidding management	Standardization of bidding management	0.317
Tender documents	Preparation of bidding documents	0.293
	Preparation of bill of quantities	0.318
Preparation of base price	Rationality of base bid preparation	0.268
Contract signing	Contract preparation	0.301
	Deviation rate of contract price	0.316
Engineering change management	Adequacy of change basis	0.266
	Changes caused by bidding documents	0.338
Claim management	Claim cost ratio	0.246
	Claim compliance	0.258
Contract management	Contract management compliance	0.205
	Validity of performance of contract documents	0.282
Project settlement	Preparation of settlement documents	0.239
	Settlement price deviation rate	0.314
	Preparation of final account documents for completion	0.336
Final account of completion	Deviation rate of final accounts	0.343

Corresponding to the parameters set in Table 2, the calculated clustering parameters are used as the numerical relationship between decoration data and threshold parameters, and the evaluation method is constructed on the basis of the numerical value.

#### 2.4 Complete the Construction of the Evaluation Method

Under the above numerical relationship, when estimating the cost of building decoration, the risk of cost estimation is easy to occur. The threshold value of cost risk is preset. Assuming that the construction period is  $t_1, t_2, \dots, t_n$ , when the construction is in period  $t_i$ , the planned cost of period  $i$  is  $C_i^*$ , and the actual cost is  $C_i$ . the cost of the prevention and control works of the completed project is calculated as  $EV(t_i)$  and the actual cost of the prevention and control works completed is  $AC(t_i)$ :

$$\begin{cases} EV(t_i) = \sum_{j=1}^i C_j^* \\ AC(t_i) = \sum_{j=1}^i C_j \end{cases} \quad (11)$$

On the basis of formula (6), the difference between the actual cost and the planned cost of the prevention and control project is calculated, and the cost deviation value  $\Delta C_c(t_i)$  is obtained:

$$\Delta C_c(t_i) = AC(t_i) - EV(t_i) = \sum_{j=1}^i C_j - \sum_{j=1}^i C_j^* \tag{12}$$

$C_j^*$  of the phased planned cost of the prevention and control project is the cost of each phase of the project determined according to the overall profit of the overall construction project. The actual cost in phase  $i$  is too high because of the risk factors of the project, which leads to the risk deviation. According to (7) and (8), the deviation threshold of cost risk is obtained,

$$R_c(t_i) = \frac{\Delta C_c(t_i)}{EV(t_i)} \tag{13}$$

Using the obtained threshold value of cost risk deviation, the entropy method is used to determine the weight value of estimated cost risk. At this time, the  $i$ -th cost deviation transmitted by the prevention and control project is defined as  $I_i$ , and  $I_i = -R_c(t_i)np_i$  is obtained, where  $p_i$  is the probability of deviation. Suppose there are  $n$  deviations at this time, and the probabilities of occurrence are  $p_i^0$  and  $i = 1, 2, \dots, n$  respectively, so the entropy of deviation is:

$$e = -\sum_{i=1}^n p_i R_c(t_i) np_i \tag{14}$$

The entropy obtained is used to determine the risk weight of cost, assuming  $x_{ij}$  is the observation value of the  $j$  index of  $I$  prevention and control project ( $i = 1, 2, \dots, n$ ). For a given  $x_j$ , the greater the difference between  $x_{ij}$  and the greater the information content contained in the index, the characteristic specific gravity of the index is calculated:  $p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}} (x_{ij} \geq 0)$ , and the entropy value of indicator  $x_{ij}$  is

$e = -k \sum_{i=1}^n p_{ij} R_c(t_i) np_{ij}$ , in which  $k > 0$ , where  $k = \frac{1}{\ln n}$  is obtained. Normalize  $e_j$ , so  $0 \leq e_j \leq 1$ . The difference coefficient of index  $x_j$  is calculated as follows:  $g_i = 1 - e_j$ , and the final cost risk weight value is obtained:

$$w_j = \frac{g_i}{\sum_{i=1}^m g_i} \tag{15}$$

where:  $w_j$  is the final calculated cost risk weight expression.  $m$  is the number of difference coefficients. Then the three-dimensional structure is used to evaluate the risk weight, and the cost estimation function is established. Before establishing the cost estimation function, the risk early warning of the prevention and control project should

be evaluated, and the three-dimensional structure chart should be used to early warn the cost risk of the prevention and control project. The three-dimensional structure chart used is shown in Fig. 5

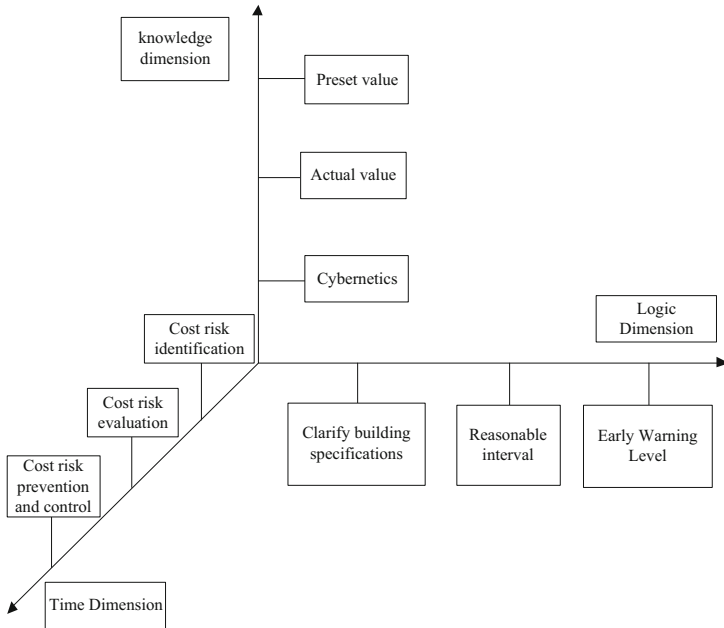


Fig. 5. Three dimensional structure of risk early warning

As shown in Fig. 4 above, cost risk is evaluated according to three dimensions: time dimension, logic dimension and knowledge dimension. In the time dimension, on the basis of equipment depreciation, we use formula (12) to identify and prevent the risk of the project. In the logical dimension, formula (13) was used to calculate the soil pollution concentration, and the logical dimension was evaluated according to the concentration. In the knowledge dimension, formula (14) is used to calculate the difference between the preset value and the actual value, so as to control the size of the preset value more strictly and realize the evaluation on the knowledge dimension.

Based on the three-dimensional evaluation results, when establishing the cost, we integrate (11), (12) and (15):

$$C = \lambda \frac{K}{R_i} \tag{16}$$

At this time, (16) is the evaluation function of cost. In actual use, according to the nature of building construction, adjust the calculation index in formula (16), calculate the weight of the index, get the weight value of cost risk, use the result of three-dimensional evaluation of cost risk weight, establish the cost estimation function, so as

to complete the construction of the cluster evaluation method of building decoration price rationality.

### 3 Simulation Experiment

The following experiments are designed to verify the effectiveness of the cluster evaluation method of building decoration cost rationality based on BIM model.

#### 3.1 Experimental Preparation

According to the arrangement of FHC decoration completion acceptance data, combined with the evaluation and weight assignment of each index, the classical domain of each index can be obtained by calculating each single index. This paper analyzes and studies the rationality of architectural decoration cost, and quantitatively and qualitatively deals with the classical domain of quantitative index of architectural decoration cost. The classical domain is shown in Table 3.

**Table 3.** Classical domain of quantitative index

Index name	Excellent	Second
Amount of contract price change	(0.7, 0.6)	(0.05, 0.1)
Contract dispute rate	(0.7, 0.6)	(0.05, 0.1)
Design change ratio	(0.7, 0.6)	(0.05, 0.1)
Change quantity and amount ratio	(0.7, 0.6)	(0.05, 0.1)
Default ratio of project payment	(0.7, 0.6)	(0.05, 0.1)
Major change ratio	(0.7, 0.6)	(0.05, 0.1)
Budget implementation rate	(0.7, 0.6)	(0.05, 0.1)
Change response rate	(0.7, 0.6)	(0.05, 0.1)
Difference from budget estimate	(0.7, 0.6)	(0.05, 0.1)
Management goal setting	(0.7, 0.6)	(0.05, 0.1)
Organization management mode	(0.7, 0.6)	(0.3, 0.5)
Clarity of responsibility division	(1, 0.7)	(0.3, 0.5)
Qualification examination	(1, 0.7)	(0.3, 0.5)
Evaluation method	(1, 0.7)	(0.3, 0.5)
List settings	(1, 0.7)	(0.3, 0.5)
Contract form	(1, 0.7)	(0.3, 0.5)
Procurement of main materials and equipment	(1, 0.7)	(0.3, 0.5)
Contract form	(1, 0.7)	(0.3, 0.5)
Clarity of contract terms	(1, 0.7)	(0.3, 0.5)



Under the quantitative and qualitative indicators set in the above table, after defining the classical domain, we can obtain the section domain of building decoration cost evaluation matter-element, and organize it into the evaluation data set as shown in Table 4.

**Table 4.** Evaluation data sets prepared

Dataset name	Dataset size/M	Time series length	Number of categories
1	2125.8	210	66
2	2377.5	170	91
3	2313.8	223	67
4	2378.4	211	88
5	2447.2	132	59
6	2479.6	245	10
7	2454.4	154	89
8	2055.7	249	87
9	2399.8	206	13
10	2487.5	142	7
11	2070.9	227	66
12	2243.3	160	94
13	2431.2	168	14
14	2032.2	179	72
15	2395.7	182	71
16	2204.8	194	75
17	2217.8	133	57
18	2020.7	191	79
19	2056.1	149	20
20	2041.4	180	80

Using the experimental data set shown in Table 4, prepare two traditional evaluation methods and the evaluation method designed in this paper for experiments, and compare the performance of the three evaluation methods.

### 3.2 Results and Analysis

Based on the above experimental preparation, under the control of the three clustering evaluation methods, the actual running time of the three clustering evaluation methods is counted according to the data set in the table. The results are shown in Table 5.

**Table 5.** Operation time of three clustering evaluation methods

Dataset name	Running time/ms		
	Traditional assessment methods 1	Traditional assessment methods 2	Method of this paper
1	36.7	28.6	11.4
2	39.8	20.9	10.6
3	40.5	29.9	10.6
4	34.9	28.6	10.7
5	35.4	20.7	10.8
6	36.9	24.2	11.3
7	31.5	24.7	11.3
8	34.4	29.5	10.1
9	36.5	23.3	11.6
10	41.9	28.2	11.3
11	38.6	24.8	11.7
12	37.9	28.2	11.6
13	32.1	24.6	10.6
14	41.3	23.7	10.2
15	41.5	22.4	10.1
16	35.9	28.6	10.9
17	35.8	24.6	10.3
18	38.9	25.3	11.2
19	39.9	21.1	11.3
20	36.5	21.9	10.9

According to the running time results shown in Table 5, the average running time of traditional evaluation method 1 is about 38S, and the actual running time is longer. The average evaluation time of traditional evaluation method 2 is about 24 s, and the actual running time is short. The average time of this method is about 10 s, compared with the two traditional evaluation methods, the time of this method is the shortest.

In the above experimental environment, the accuracy of three reasonable clustering evaluation methods is defined:

$$C = \lambda \frac{K}{R_i} \quad (17)$$

Among them,  $C$  is the accuracy of evaluation,  $\lambda$  is candidate parameter,  $K$  is evaluation data set, and  $R_i$  is cluster parameter. Under the control of the above numerical relationship, the accuracy results of the final three evaluation methods are shown in Table 6.

**Table 6.** Accuracy results of three evaluation methods

Dataset name	Accuracy		
	Traditional assessment methods 1	Traditional assessment methods 2	Method of this paper
1	0.5294	0.6955	0.9062
2	0.5673	0.6951	0.9135
3	0.4004	0.7676	0.9337
4	0.4303	0.7465	0.9076
5	0.5209	0.6134	0.9315
6	0.5411	0.6154	0.9083
7	0.5879	0.7711	0.9073
8	0.5194	0.7981	0.9399
9	0.5119	0.7888	0.9378
10	0.5318	0.6187	0.9359
11	0.5689	0.7849	0.9383
12	0.4459	0.6204	0.9388
13	0.4999	0.7934	0.9469
14	0.4279	0.6273	0.9093
15	0.5423	0.6916	0.9141
16	0.4417	0.6137	0.9223
17	0.5281	0.7162	0.9151
18	0.4344	0.6566	0.9141
19	0.5499	0.6299	0.9203
20	0.4949	0.6456	0.9394

Under the control of three evaluation methods, the accuracy relationship constructed above is taken as the accuracy result. According to the numerical results in Table 6, it can be seen that the accuracy of traditional evaluation method 1 is about 0.5, and the actual evaluation accuracy is poor. The accuracy of traditional evaluation method 2 is about 0.7, and the accuracy of the actual evaluation results is higher. The accuracy of this method is about 0.9, and the accuracy of this method is the best compared with the two traditional methods.

The above experimental environment is kept unchanged, and the quantitative classical field in the above table is different gradients. The calculation process can be expressed as follows:

$$G = \frac{T - T'}{H_0 - h} \quad (18)$$

Among them,  $T$  is the classical domain acquisition cycle,  $T'$  is the ideal evaluation cycle,  $H_0$  is the domain value difference, and  $h$  is the number of indicators. According to the above calculation formula, a total of five reasonable gradients are collected. Under the control of the reasonable gradient, three evaluation methods are used for

processing. The more data points that can be evaluated in the corresponding gradient, the greater the efficiency of the evaluation method is. Finally, the evaluation efficiency results of the three evaluation methods are shown in Fig. 6.

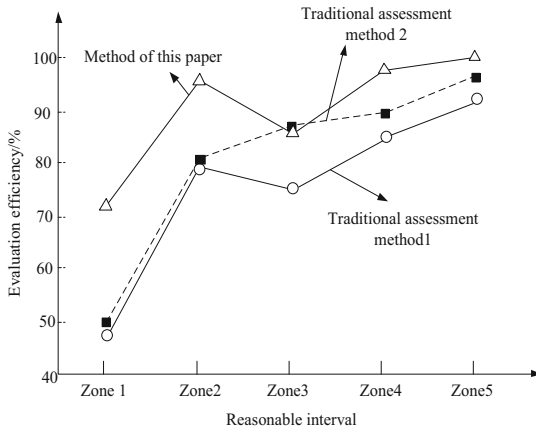


Fig. 6. Efficiency results of three evaluation methods

According to the efficiency results of the evaluation method shown in Fig. 6, after dividing the reasonable gradient of cost, the average efficiency of traditional evaluation method 1 is about 78%, that of traditional evaluation method 2 is about 82%, and that of this method is about 93%, which proves that the efficiency of this method is the highest compared with the two traditional evaluation methods.

## 4 Conclusion

Through the analysis of BIM model on the current situation of project cost management, it can be found that the owner's evaluation of project cost management ability stays on empiricism and lacks systematic and scientific methods. Therefore, this paper constructs the construction cost cluster evaluation method, which can evaluate the rationality of the construction decoration cost, evaluate the maturity of the cost management in each stage, and finally get the maturity level of the project cost management, and improve it pertinently, so as to provide a more scientific evaluation method for the construction decoration cost. The current construction project cost prediction model has problems such as complex modeling and ignoring regional differences. The construction cost per unit area is used to cluster the residential construction projects in 30 areas, and the 30 areas are divided into four categories. The next step of the research direction can be to add the area category as an influencing factor to the linear regression model, and obtain the prediction model of the residential building project cost through principal component regression analysis.

## Fund Projects

1. Guangxi Vocational Education Teaching Reform Research Project in 2020 “Research and Practice on the Practice Teaching System of “Three Courses and Four Abilities” in Higher Vocational Architecture Majors with Progressive Ability” (GXGZJG2020B122)
2. In 2020, Guangxi Gao School's young and middle-aged teachers' scientific research ability improvement project “Research and Practice of Prefabricated Buildings Based on BIM Technology in Underdeveloped Areas” (2020KY45013).

## References

1. Wang, A., Su, M., Sun, S., et al.: Evaluation of BIM application capability of prefabricated buildings based on life cycle theory. *J. Civil Eng. Manage.* **37**(02), 27–33+40 (2020)
2. Dong, N., Hu, L., Zou, Z., et al.: Application benefit evaluation of BIM based on C-OWA operator and grey clustering. *Sci. Technol. Manage. Res.* **39**(04), 48–544 (2019)
3. Wang, W., Ma, B., Li, Q., et al.: Clustering of BIM components based on similarity measurement of attributes. *J. Graph.* **41**(02), 304–312 (2020)
4. Li, P., Xu, J., Duan, Y.: Evaluation method of property right of building information modeling in construction enterprises. *Sci. Technol. Manage. Res.* **39**(01), 183–189 (2019)
5. Sun, S., Sun, B., Wu, J.: Evaluation on BIM-based synergy degree in design stage of prefabricated buildings. *Yangtze River* **51**(04), 218–225 (2020)
6. Han, F., Lin, S.: Energy-saving mode assessment method for hydropower building construction based on building information modeling technology. *Sci. Technol. Eng.* **19**(24), 231–236 (2019)
7. Liu, S., Liu, D., Khan, M., Ding, W.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>
8. Liu, S., Liu, X., Wang, S., Muhammad, K.: Fuzzy-Aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
9. Bai, J., Lyu, X., Sun, P.: Rapid modeling method of urban rail transit based on BIM API. *Urban Mass Transit* **22**(08), 170–173 (2019)
10. Zhao, W.: Dynamic integration mechanism of building information modeling technology and low carbon green building evaluation index system. *Sci. Technol. Eng.* **19**(03), 196–201 (2019)
11. Liu, S., Sun, G., Fu, W.: e-Learning, e-Education, and Online Training, pp. 1–386. Springer International Publishing, USA
12. Hu, Y., Pei, L., Han, C.: A seismic performance evaluation method for masonry structures based on BIM. *China Earthquake Eng. J.* **41**(01), 221–226 (2019)
13. Chen, H., Xu, G., Wu, X., et al.: Research on information design and green degree evaluation of green building based on BIM. *Archit. Technol.* **50**(08), 996–1000 (2019)



# Power Frequency Control Method of Electrical Equipment Under Internet of Things Technology

Hong-yu Huang<sup>(✉)</sup> and Qing-huan Qin

Guangxi Modern Polytechnic College, Hechi 547000, China  
huanghongyu6542@163.com

**Abstract.** The power frequency of electrical equipment is related to the safety of the equipment, but the traditional control method is difficult to guarantee the safety of the electrical equipment because of the small difference between the activity characteristics of the power information extracted by the traditional control method. Therefore, this study proposes a power frequency control method for electrical equipment based on Internet of Things technology. Firstly, power supply information of electrical equipment is collected based on the Internet of Things technology, and then a collaborative model is established to extract the frequency fluctuation characteristics of power supply. Based on this, the frequency fluctuation range is located, the node information is compressed sensing on the basis of two feature data processing, and then chaotic algorithm is used to control the frequency fluctuation trajectory of the power supply. Four groups of variables were set to test the safety performance guarantee effect of the control method. Compared with the traditional control method, the safety factor of the electrical equipment is 7.96% higher than that of the traditional control method. It can be seen that the method in this paper can effectively improve the use safety of electrical equipment.

**Keywords:** Internet of Things technology · Electrical equipment · Power frequency control · Safety factor

## 1 Introduction

With the rise of network science and technology, foreign countries first began to study signal transmission control theory through mathematical abstract hypotheses in 1736. After that, for nearly two centuries, people have been combining mathematical abstract hypotheses to control the synchronous transmission of signals. The domestic research on this work started late, but with the continuous deepening and innovation of various science and technology, various synchronous control strategies or methods have different research progress. Among them, in reference [1], a 5 kHz high-frequency ice-melting excitation power supply is designed to achieve large-capacity output of the system through parallel inverter modules, and then the drive signal of the inverter switch is generated by carrier phase shift technology to reduce the working frequency of the switch, so as to improve the control effect of the output frequency of the power supply. In Reference [2], on the basis of determining the power frequency control range

of the power electronic interface, the equivalent inertia and the difference adjustment coefficient of the power system are calculated, and then the frequency control of the power system is realized through the frequency support between multi-synchronous networks.

On the basis of the above research, this paper combines the existing frequency control technology of different objects to carry out optimization research on the power frequency control of electrical equipment. The main work of this paper is as follows:

- (1) Collect power supply information of electrical equipment based on Internet of Things technology and establish a collaboration model, so as to extract frequency fluctuation characteristics of power supply.
- (2) By positioning the frequency fluctuation range, and on the basis of two characteristic data processing, compressed sensing node information is used, and then chaos algorithm is used to control the power frequency fluctuation trajectory, so as to realize the effective control of power frequency.

## 2 Method Design

### 2.1 Collect Electrical Equipment Power Information Based on the Internet of Things Technology

The Internet of Things technology includes information collection technology, wireless network communication technology, wireless positioning technology, and smart technology. The Internet of Things technology is used to monitor the power supply of electrical equipment. The Internet of Things RFID sensor technology uses TDOA (Time Difference of Arrival Algorithm) to obtain the activity status of electrical equipment in different cycles according to the hyperbola [3, 4]. Assuming that  $A$ ,  $B$ ,  $C$ , and  $D$  represent the position of the RFID reader, it is known that the position coordinates are statically fixed. Set  $E$  to represent the position of the node that needs to be located. The TDOA algorithm uses points  $AB$  and  $CD$  to create a hyperbola. When point  $P$  is within the radio frequency range of the radio frequency reader, the following formula is used to calculate the position coordinates of point  $P$ .

$$\begin{cases} W_{AB} = t_1 v = |PA - PB| \\ W_{CD} = t_2 v = |PD - PC| \end{cases} \quad (1)$$

In the formula:  $W_{AB}$  and  $W_{CD}$  respectively represent the distance difference between point  $P$  to point  $A$  and point  $B$ ;  $t_1$  and  $t_2$  represent the time difference between the radio frequency signal at point  $P$  and point  $A$  and  $B$  respectively;  $v$  represents the signal transmission speed;  $PA$ ,  $PB$ ,  $PD$ ,  $PC$  represent the distance between point  $P$  and the RFID reader. Bring the calculation result of the above formula into the hyperbolic formula to calculate the position coordinates of the  $P$  point. The equation is:

$$\begin{cases} R_{AB} = \sqrt{(X_A - X)^2 + (Y_A - Y)^2} - \sqrt{(X_B - X)^2 + (Y_B - Y)^2} \\ R_{CD} = \sqrt{(X_C - X)^2 + (Y_C - Y)^2} - \sqrt{(X_D - X)^2 + (Y_D - Y)^2} \end{cases} \quad (2)$$

In the formula:  $(X_A, Y_A)$ ,  $(X_B, Y_B)$ ,  $(X_C, Y_C)$ , and  $(X_D, Y_D)$  represent the coordinates of points  $A$ ,  $B$ ,  $C$ , and  $D$  respectively, and  $(X, Y)$  is the coordinates of point  $P$  for calculation. Solve the above formula and calculate the specific  $(X, Y)$  value [5, 6]. After the Internet of Things technology collects electrical equipment power information, it uploads the information to the control platform, and uses the platform's correlation analysis and linkage operation functions to transform the island information that was originally in an isolated state into linkage information.

### 2.2 Establish a Collaborative Model to Extract the Characteristics of Power Frequency Fluctuations

On the premise of the integration of the digital control platform and the Internet of Things technology, according to the collaborative mechanism of the Internet of Things and electrical equipment, a collaborative model is established to extract the characteristics of power frequency fluctuations. Assuming that the operational risk of the control platform is neutral, and the various data in the platform can be regarded as a single mechanism of approximate project discrete changes, a two-state model of binomial distribution conforming to a single period is created. In the basic management module of electrical equipment, considering the flexibility of the control program, it is assumed that the decision maker adjusts the original operation plan according to the actual operating conditions of the electrical equipment within a certain period of time [7]. Assuming that the value-added amount is  $M$ , the risk probability when the platform is used at this time is:

$$o = \frac{t_i}{t_0} \times R(n + \varepsilon) \quad (3)$$

In the formula:  $o$  represents the risk probability;  $t_i$  represents the power data adjustment time in the  $i$  period;  $t_0$  represents the initial time of the dynamic change of the power data;  $n$  represents the original data volume;  $\varepsilon$  represents the increased or decreased data volume after adjustment using  $t_1, t_2, t_3$  respectively represents different equipment operating periods. According to the change probability of power-related data, a collaborative analysis model is established:

$$K = e^{-k} [o \times (\varphi a + \delta) + (1 - o) \times (\varphi b + \delta)] \quad (4)$$

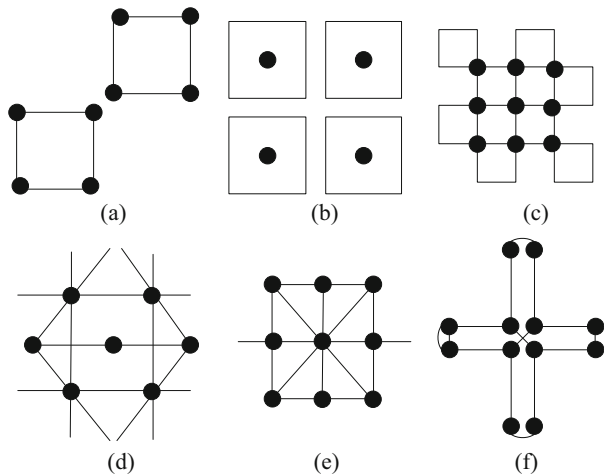
In the formula:  $e$  represents the risk identification coefficient;  $k$  represents the risk indicator;  $\varphi$  represents the volatility value of the platform;  $a$  represents the rate of increase in risk;  $b$  represents the rate of decrease in risk. Use the model to mine and analyze the data characteristics of the electrical equipment power information, compare the results with the actual data characteristics, and find out the related parts between the



two. Using grid fast density clustering algorithm, the electrical equipment power data information is divided into several adjacent intervals according to the dimensions, and a set composed of grid cells is established. A dynamic two-dimensional array is used to access the entire data set. Unordered data sets are transposed into an ordered structure in a specific space to achieve clustering of data. The data space in the data set is divided by the method of equal width division [8]. Assuming that the two-dimensional data object is  $u$ , and its neighborhood is centered at  $u$  and  $r$  is the radius, the collection of all data objects forms a single data set. Therefore, it is appropriate to set the width of the grid to be a value. Suppose the two-dimensional data set is  $Z$ , its attribute is  $(U, V)$ , the values of  $U$  and  $V$  dimensions are  $(u_{\min}, u_{\max})$ ,  $(v_{\min}, v_{\max})$ , and  $r$  is the grid width for division, and the specific number of grids obtained is:

$$\begin{cases} U = \left( \frac{u_{\max} - u_{\min}}{r} \right) \\ V = \left( \frac{v_{\max} - v_{\min}}{r} \right) \end{cases} \quad (5)$$

The number of grids in each dimension takes an integer value upwards. Read the data set  $Z$  and divide it into a grid of  $U \times V$ . After the grid unit corresponding to each list is initialized, the power supply data of electrical equipment is read into the memory, and the attribute characteristics are calculated by the allocation function according to the actual spatial position, and then placed into the corresponding grid. At this time, there are many types of data features mined. The power data feature types are shown in Fig. 1.



**Fig. 1.** Schematic diagram of power supply data feature types

From the above figure, it can be seen that there are many hidden layers of data and complex feature attributes. Therefore, the data output results in each hidden layer are calculated according to formula (4) and formula (5):

$$g(u, v) = K \left[ \sum_{i=1}^n \delta_{ij}(U \times V) - \theta_j \right] \quad (6)$$

In the formula:  $g(u, v)$  represents the amount of data output in the hidden layer within the dimensional value range of the data attribute;  $\delta_{ij}$  represents the weight coefficient between the input layer  $i$  and the hidden layer  $j$ ;  $\theta_j$  represents the threshold in the hidden layer  $j$ . When the difference between the calculation result and the actual data value is greater than 0.3, it means that the data has an error and needs to be recalculated; when the calculated result is similar to the actual power data under the Internet of Things, it means that the amount of data obtained by the model is credible. Yes, the data characteristics at this time are consistent with the actual situation.

### 2.3 Positioning Frequency Fluctuation Range Compressed Sensing Node Information

When users access the platform to obtain power information of electrical equipment, they are affected by network interference signals. Although the power information obtained has retrieval characteristics, it is scattered in the entire retrieval set. Considering the relevance and interactivity between network data transmission nodes, The network traffic transmission process is regarded as a movement process. The number of anchor nodes of the platform is known by default. When locating the range of feature data in the collection, assuming that the number, location and network topology of anchor nodes are fixed, set one of the target nodes to 0 and use  $n$  mobile anchor nodes to calculate location. Set the number of anchor nodes to  $1, 2, \dots, n$ , one of the random nodes is  $i$ , the position is  $q_i$ , there is  $q_i = \{q_{i,1}, q_{i,1}, \dots, q_{i,m}\}$ , where  $m$  represents the dimension of the network space, and  $n \geq m + 1$ . Let the target node to be located be  $q_0$ , and calculate the distance between this node and other nodes. Using the idea of non-adjacent subtraction of the Internet of Things technology to eliminate the quadratic term in the multilateral positioning process. Assume that anchor node  $i$  needs to meet the following constraints:

$$l_{0,i} = \sqrt{\sum_{j=1}^m (q_{0,j} - q_{i,j})^2} \quad (7)$$

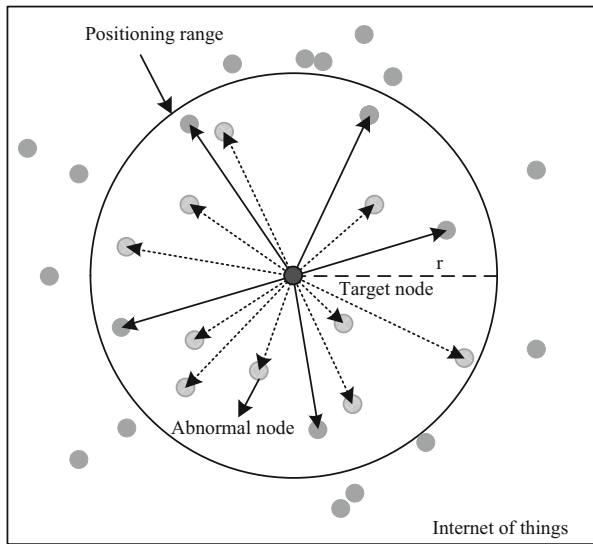
In the formula:  $q_{0,j}$  represents the unknown variable to be determined. In order to facilitate the safe use of the Internet of Things technology, the above formula is transformed to obtain:

$$l_{0,j}^2 - \sum_{j=1}^m q_{i,j}^2 = \sum_{j=1}^m q_{0,j}^2 - 2 \sum_{j=1}^m q_{0,j} q_{i,j} \quad (8)$$

According to the above conversion results, the  $n$  formula is used to subtract the  $i$  formula to eliminate the quadratic term on the left side of formula (8) to obtain a linear equation  $Xq_0^T = Y$ . According to the definition of  $X$  and  $Y$  in the formula, the linear equation estimation result:

$$\hat{q}_0^T = (X^T X)^{-1} X^T Y \quad (9)$$

The above results are the farthest distance of abnormal nodes in the Internet of Things with a target node as the center after calculation. Repeat the above steps to get the mining range consisting of all characteristic data in the set. The mining scope is shown in Fig. 2.



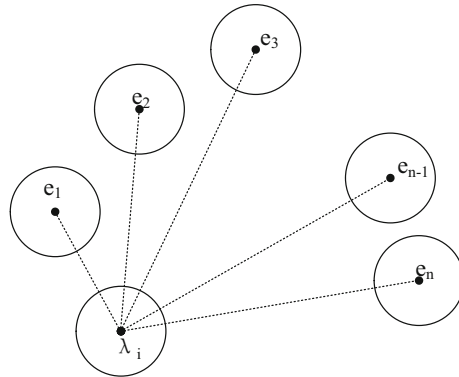
**Fig. 2.** Schematic diagram of positioning range

After the Internet of Things technology fully covers and senses the information link, information exchange and complete coverage as shown in Fig. 2 can be realized according to the changes of data characteristics of each node in the set. Then, through feature recognition of the nodes in the whole set, the range of the required power characteristic information node is located. After determining the range, a feature extraction method based on the sum of the distance between cluster centers is used to extract the abnormal node features due to the link failure. This process divides the obtained range into a training set and a test set, and then extracts  $k$  cluster centers from the training set, and then transforms the original  $m$  dimensional data set to obtain a brand-new  $k$  dimensional data set. The process consists of two main stages. The first stage is to cluster after the training set is obtained through the above method, and then extract the cluster center; the second stage is to calculate the sum of distances between the target node and each node again, so as to construct a new training set training data.

Suppose that after the  $k$  dimensional data set  $G$  is divided, the training set and the test set obtained are  $G_1$  and  $G_2$  respectively, and  $k$  unconnected clusters and cluster centers in  $G_1$  are aggregated and extracted. Since the behavioral characteristics between nodes in the first stage are relatively similar, the data set needs to be divided into multiple clusters [9, 10]. Assuming that the data sample in  $G_1$  is  $\lambda_i$ , calculate the distance between  $\lambda_i$  and all clusters, and obtain the closest cluster  $E_j$ , repeat this step to obtain  $k$  cluster centers  $e_1, e_2, \dots, e_k$ . The objective function at this time is:

$$\tau = \sum_{j=1}^k \sum_{i=1}^n \left\| \lambda_i^{(j)} - e_j \right\|^2 \tag{10}$$

In the formula:  $V$  is the distance measurement between the data sample  $\lambda_i^{(j)}$  and the cluster center  $e_j$  of the cluster where it is located;  $n < k$  represents the total amount of  $k$  dimensional data used. Through the above steps,  $k$  cluster centers in  $G_1$  are obtained. Use  $G_1$  and the cluster center to generate a  $k$  dimensional data set  $G'_1$ , where each data sample is composed of the sum of  $k$  distances. According to the above analysis, the cluster center  $e_1, e_2, \dots, e_k$  is used to transform  $G_2$  into a new data set  $G'_2$ , and the classification algorithm is used to classify the result of  $G'_2$ . Given the distance within the positioning range, for any two data existing in the set, the relationship between the abnormal node and the abnormal edge node exists as shown in Fig. 3.



**Fig. 3.** Association between abnormal nodes and abnormal edge nodes

According to the mechanism shown in the figure, calculate the Euclidean distance between them, the formula is as follows:

$$dis(\lambda_i, e_j) = \sqrt{(\lambda_1 - e_1)^2 + \dots + (\lambda_n - e_n)^2} \tag{11}$$

In the formula:  $n$  represents the number of abnormal nodes. By repeating the above calculation process, a new data set  $G'_2$  is obtained, and then node classification is performed to find out the strong and weak associations between multiple nodes and the

target node, and obtain the characteristics of all power data nodes with abnormal problems within the positioning range.

Among all the nodes obtained by positioning, the abnormal power information nodes are not evenly distributed. A distributed detection algorithm needs to be used to realize compressed sensing of all abnormal nodes, so as to realize the cyclic detection of retrieved data. The algorithm divides the entire positioning range into several subsets, and performs compressed sensing tasks in each subset. After the detection algorithm completes the loop detection work, the detection parameters are adjusted according to the detection results, and the corresponding detection matrix is adjusted at the same time [11]. Suppose the number of nodes in the positioning range is  $S$ , and  $x(s)$ ,  $s = 1, 2, \dots, S$  represents the discrete data vector. The detection algorithm performs a weighted tracking task for each location area, and uses the following formula to determine whether the data of each node exists because the link causes the characteristic value of the power supply to be abnormal:

$$F[x'(i)] = \begin{cases} \varphi_1 & \text{if } x'(i) > \beta, \forall i = 1, 2, \dots, S \\ \varphi_0 & \text{otherwise} \end{cases} \quad (12)$$

In the formula:  $x'(i)$  represents the  $i$  recovered data obtained after the algorithm weighted tracking;  $\varphi_1$  represents the abnormality of the characteristic data node;  $\varphi_0$  represents the abnormality of the characteristic data node;  $\beta$  represents the abnormal threshold of the node. Compare the calculation result with the actual situation, and adjust the regional weight of each positioning range. Summarize the detection situation of the entire collection, locate the detection result position from a global angle, and then adjust the detection matrix in the corresponding area. Here we assume that the set is divided into  $v$  sub-intervals, from which the detection model obtains  $b$  node information, and adjusts the number of detection nodes according to the following formula:

$$b_{ij} = b \times \frac{H_{ij}}{\sum_{i=1}^v H_{ij}} \quad (13)$$

In the formula:  $H_{ij}$  and  $H_{ij}$  respectively represent the number of data detections and the number of abnormal nodes found in the  $i$  subset during the  $j$  detection of the algorithm. The algorithm compresses and senses abnormal node information, so that the obtained electrical equipment power information characteristics are consistent with the retrieval target, and at the same time, the abnormal characteristic data of its own is removed to ensure that the control content is consistent with the target content.

#### 2.4 Chaos Algorithm to Control Power Supply Frequency Fluctuation Trajectory

According to the characteristic parameters obtained above, the chaotic algorithm is used to track the fluctuation trajectory of the power frequency of the electrical equipment. The particle swarm optimization algorithm in the chaos algorithm finds the optimal solution for the particles in space through the way birds find food. Each particle

is used as a dimensional variable. Therefore, the general calculation equation of the chaotic particle algorithm is:

$$\begin{cases} \min L(\sigma) = L(\sigma_1, \sigma_2, \dots, \sigma_n) \\ \sigma_{i \min} < \sigma_i < \sigma_{i \max}, i = 1, 2, \dots, P \end{cases} \quad (14)$$

In the formula:  $\sigma_i$  represents the dimensional variable of the  $i$  particle;  $P$  represents the particle set.  $L(\sigma)$  represents the result of particle swarm algorithm. Suppose that the  $i$  particle is denoted by  $T_i$ , and its most suitable value in the iterative process is  $T_0$  by default. In the iterative process, the best particle produced in each generation is  $J_0$ , then the iterative formula of the chaotic particle swarm algorithm is:

$$\alpha_{ij}^{e+1} = \delta \alpha_{ij}^e + \kappa_1 \zeta_1 (\alpha_{ij}^e - \sigma_{ij}^e) + \kappa_2 \zeta_2 (J_{ij}^e - \sigma_{ij}^e) \quad (15)$$

In the formula:  $\alpha_{ij}^e$  represents the  $j$  dimensional velocity of the  $i$  particle;  $\delta$  represents the inertia factor;  $\kappa_1$  and  $\kappa_2$  are calculated by the matrix;  $\zeta_1$  and  $\zeta_2$  represent random numbers in the range;  $\sigma_{ij}^e$  represents the  $j$  dimension of the  $i$  particle variable. In the algorithm used in traditional control methods, as long as a particle is found closest to the optimal particle, the particle is regarded as the optimal particle. No matter whether there is a more suitable optimal solution later, the algorithm does not continue to iterate. The chaotic particle algorithm used this time uses the chaotic mapping method to map the particles into the solution space, while maintaining the diversity of the particles, it breaks through the defect of local optimization [12]. The Logistic mapping of the chaotic algorithm is the basic performance of a typical mapping. The initial equation of the mapping is:

$$f_{i+1} = \chi f_i (1 - f_i) \quad (16)$$

In the formula:  $\chi$  represents a constant with a range between [3.57,4];  $f_i$  represents the range of a chaotic sequence under Logistic mapping, with a value between (0,1). The logistic equation satisfies the traversal requirements of the chaotic algorithm. At this time, the power source signal source in the electrical equipment moves through all reachable and non-repetitive state points in the chaotic domain. Knowing the existence range of the chaotic sequence  $f_i$ , the variable value  $\sigma_i$  of the chaotic algorithm is mapped from the separate solution space to the chaotic sequence. Therefore, the calculation result of the comprehensive formula (14–16) is obtained, and the mapping equation of the chaotic variable is obtained as:

$$f_i = \frac{\sigma_i - \sigma_{i \min}}{\sigma_{i \max} - \sigma_{i \min}} \quad (17)$$

In the formula:  $\sigma_{i \min}$  and  $\sigma_{i \max}$  respectively represent the minimum and maximum values of the chaotic variable  $\sigma_i$ . According to the above calculation, the chaotic sequence mapping result is obtained and mapped to the variable value solution space of the algorithm, then:

$$\sigma_i = f_i(\sigma_{i\max} - \sigma_{i\min}) + \sigma_{i\min} \tag{18}$$

Logistic mapping is used to obtain the parameter values of the chaotic sequence particles, and the fluctuation trajectory of the power signal is tracked according to the results. This research requires the control of the power frequency of electrical equipment. Therefore, combining the premise that the sum of the three-phase current and the sum of the three-phase error current are zero, we get:

$$\Delta I_{ab} = -(\Delta I_{bc} + \Delta I_{ca}) \tag{19}$$

In the formula:  $\Delta I_{ab}$ ,  $\Delta I_{bc}$ , and  $\Delta I_{ca}$  respectively represent the error currents between phases  $ab$ , phase  $bc$ , and phase  $ca$ . Therefore, if the error current between the two phases can be controlled, and the phase is opposite, the third phase error current can be limited to the minimum range, thereby reducing the power loss of the electrical equipment. According to the variation of the error current, the controlled phase-to-phase error current requires fixed-frequency control. A standard reference frequency pulse is used as the control reference signal, and the phase of each phase switch is aligned with the reference signal to ensure that the phase switches can operate synchronously. In the initial stage of control, the proposed control method stabilizes the switching frequency in a fixed frequency range, and then fixes the phase difference between the switching phase and the standard pulse signal. When adjusting the frequency, a phase shift is used as the compensation for the hysteresis width [13]. By default, in the last switching cycle, the time difference between the midpoint of the 0-value signal and the pulse signal is  $\Delta c$ , then it is derived:

$$C = \left( \Delta c + \frac{R}{2R/C_2} \right) + \frac{R+r}{2R/C_1} + \frac{r}{2R/C_2} \tag{20}$$

In the formula:  $C$  represents the expected target of the switching signal level action time in the next cycle;  $C_1$  and  $C_2$  represent the high level and low level action time of the switching signal in the next cycle respectively[14];  $R$  represents the current hysteresis width. Taking into account the compensation of the error current variation, assuming that the compensation is  $\Delta\phi_1$  and  $\Delta\phi_2$  respectively, the revised calculation result of formula (20) is:

$$C' = \left( \Delta c + \frac{R}{2R/C_2} \right) + \frac{R+r}{2R/C_1 + \Delta\phi_1} + \frac{r}{2R/C_2 + \Delta\phi_2} \tag{21}$$

In the formula:  $C'$  is the revised expected target under constant-frequency hysteresis current control [15]. According to the above process, the control of the power frequency of electrical equipment under the chaotic algorithm is realized.

### 3 Experimental Study

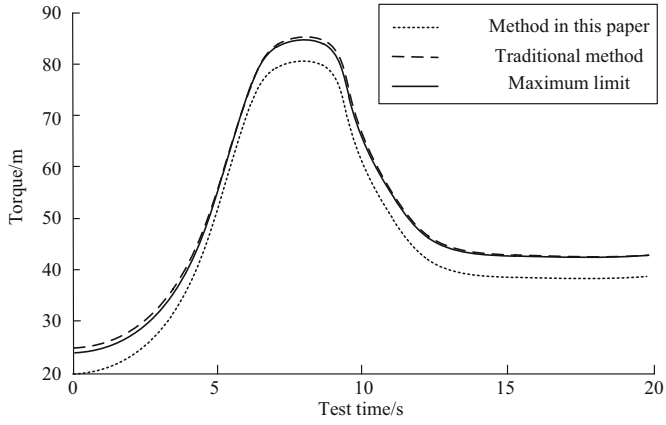
In order to further verify the control effect of the research method, the control method is applied to a simulation test environment, and a group of traditional control methods are used as a control group to compare the control differences between different methods. The experiment selects a certain electrical equipment in the power plant as the test object, uses simulation software to build a test environment, and sets the simulation experiment parameters shown in Table 1.

**Table 1.** Simulation experiment parameters

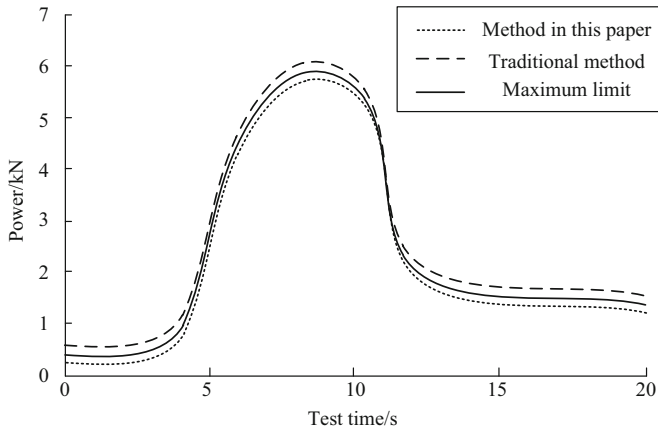
Serial number	Simulation parameters	Analog value
1	Air density	1.288 kg/m <sup>3</sup>
2	Blade radius	1.85 m
3	Rated wind speed	12.8 m/s
4	Rated power	5.6 KW
5	Tip speed ratio	8.4
6	Wind energy coefficient	0.5
7	Running constant	0.0365
8	Moment of inertia	1.15 kg/m <sup>2</sup>
9	Friction damping	0.05 N·s/rad
10	Number of pole pairs	5
11	Generator rated power	5.5 KW
12	Generator rated speed	585 r/min
13	Stator inductance	8.9 mH
14	Stator resistance	0.33 Ω
15	Main flux	0.35 × 10 <sup>-3</sup> Wb
16	Sliding mode controller parameters	0.01

According to the simulation parameters shown in Table 1, a test environment was established to test the control performance of the two methods. Figure 4 shows the changes in the electromagnetic torque, electromagnetic power, rotor angular speed and control signal of the generator's electromagnetic torque, electromagnetic power, rotor angular velocity and control signal after controlling the power frequency of the electrical equipment under the application of the two methods.

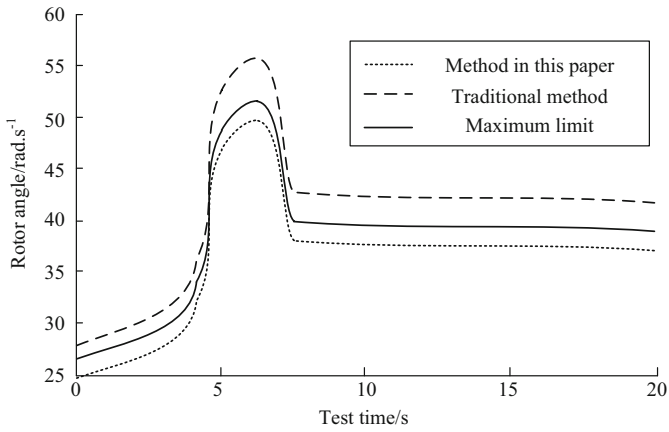




(a) Generator electromagnetic torque

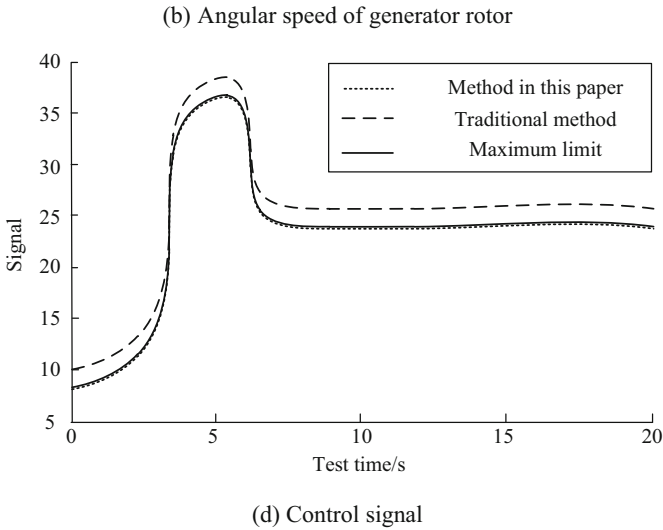


(b) Generator electromagnetic power



(c) Angular speed of generator rotor

**Fig. 4.** Control effect comparison test result



**Fig. 4.** (continued)

From the observation of the curves of the above four groups of test results, it can be seen that, on the basis of controlling the power supply frequency, the method in this paper accurately allocates data to the control center, so that the electromagnetic torque, electromagnetic power, rotor angular velocity and control signal of the generator of the entire electrical equipment are constrained within the limit value. However, the traditional method does not restrict the four parameters, which leads to the data overrun and brings great security risks for the use of electrical equipment.

In order to ensure the reliability of the test results, 10 rounds of testing were conducted for each group of tests, and AHP was used to evaluate the use safety of electrical equipment under four conditions. The evaluation results are shown in Table 2, Table 3, Table 4 and Table 5.

**Table 2.** Safety evaluation results for the use of electrical equipment (1)

Number of test rounds	Method in this paper	Traditional method
Round 1	0.9545	0.8863
Round 2	0.9536	0.8925
Round 3	0.9627	0.8842
Round 4	0.9548	0.8991
Round 5	0.9722	0.9036
Round 6	0.9716	0.8817
Round 7	0.9664	0.8988
Round 8	0.9582	0.8991
Round 9	0.9525	0.8865
Round 10	0.9817	0.8876
Average	0.9628	0.8919

**Table 3.** Safety evaluation results for the use of electrical equipment (2)

Number of test rounds	Method in this paper	Traditional method
Round 1	0.9721	0.8841
Round 2	0.9548	0.8743
Round 3	0.9725	0.8792
Round 4	0.9704	0.8786
Round 5	0.9636	0.8865
Round 6	0.9628	0.8892
Round 7	0.9714	0.8743
Round 8	0.9594	0.8907
Round 9	0.9545	0.8855
Round 10	0.9636	0.8628
Average	0.9645	0.8805

**Table 4.** Safety evaluation results of the use of electrical equipment (3)

Number of test rounds	Method in this paper	Traditional method
Round 1	0.9662	0.8584
Round 2	0.9701	0.8636
Round 3	0.9625	0.8427
Round 4	0.9488	0.8448
Round 5	0.9529	0.8533
Round 6	0.9443	0.8525
Round 7	0.9421	0.8672
Round 8	0.9557	0.8564
Round 9	0.9543	0.8565
Round 10	0.9625	0.8566
Average	0.9559	0.8552

**Table 5.** Safety evaluation results of use of electrical equipment (4)

Number of test rounds	Method in this paper	Traditional method
Round 1	0.9439	0.8763
Round 2	0.9488	0.8929
Round 3	0.9443	0.8876
Round 4	0.9447	0.8764
Round 5	0.9525	0.8852
Round 6	0.9546	0.8927
Round 7	0.9499	0.8914
Round 8	0.9476	0.8865
Round 9	0.9528	0.8829
Round 10	0.9537	0.8901
Average	0.9493	0.8862

This paper mainly analyzes the average safety factors in the above tables. Through data comparison, it can be found that the average safety factor of electrical equipment is 0.9581 after the application of the method in this paper. After the traditional method is applied, the average safety factor of electrical equipment is 0.8785, and there is a significant difference between the two. It can be seen that the method presented in this paper has a better control effect on the power frequency of the equipment and can effectively improve the safety of the equipment.

## 4 Conclusion

In this paper, a new power frequency control method for electrical equipment is proposed by combining Internet of Things technology, collaboration model, detection algorithm and chaos algorithm. This method firstly uses Internet of Things technology to collect power supply information of electrical equipment, and accurately extracts the frequency fluctuation characteristics of power supply, which lays a good foundation for subsequent control process. Then, by locating the frequency fluctuation range, accurate compressed sensing node information is obtained on the basis of two characteristic data processing. Then, chaotic algorithm is used to control the fluctuation trajectory of power frequency. The method has achieved success in the absorption stage, but it involves a lot of calculation process, so the efficiency of the method may be inferior to the traditional method. In the future, the calculation steps need to be simplified to improve the efficiency of the method.

**Fund Projects.** 2016 Guangxi Vocational Education Teaching Reform Project: Study on the Connection of Project Design between Specialized Courses in “Project Teaching” of Applied Electronic Technology Major in Higher Vocational Education, (GXGZJG2016B003).

## References

1. Wang, Y., Zhou, Y., Yuan, X.: Study on high frequency ice-melting excitation source and control strategy. *Electr. Meas. Instrument.* **56**(10), 140–146 (2019)
2. Zhang, X., Wang, J.: Research on power system frequency control method considering power electronic interface. *Electron. Test* **15**(18), 30–31+83 (2020)
3. Xu, Z., Qin, H.: Design of intelligent light pole system based on IoT technology. *Transd. Microsyst. Technol.* **39**(06), 77–78+82 (2020)
4. Wu, J., Zhao, X., Zhao, J.: Modulation and demodulation of LoRa internet of things technology. *Comput. Eng. Design* **40**(03), 617–622 (2019)
5. Zheng, F., Dai, M., Chen, H.: Research on power cooperative control of multi-inverters for Power-Supply of microgrid. *Comput. Simul.* **36**(03), 137–141 (2019)
6. Sun, B., Tang, Y., Ye, L., et al.: Integrated frequency control strategy for wind power cluster with multiple temporal-spatial scale coordination based on H-DMPC. *Proc. CSEE* **39**(01), 155–167+330 (2019)
7. Liu, S., Sun, G., Fu, W.: *e-Learning, e-Education, and Online Training*, pp. 1–386. Springer, Boston (2018). <https://doi.org/10.1007/978-3-030-63952-5>

8. Chen, G., Li, Z., Guo, Y., et al.: Research on simulation of load frequency control for interconnected power grid based on Simulink. *Exp. Technol. Manag.* **36**(01), 124–129 (2019)
9. Liu, S.: Introduction of key problems in Long-Distance learning and training. *Mob. Netw. Appl.* **24**(01), 1–4 (2019)
10. Zhang, B., Tan, W., Li, J.: Tuning of linear active disturbance rejection control for load frequency control systems with hydro turbines. *Electr. Mach. Control* **23**(01), 117–124 (2019)
11. Li, X., Jia, H., Mu, Y., et al.: Coordinated frequency control based on electric vehicles and heat pumps considering time-delay. *Electr. Power Autom. Equip.* **40**(04), 88–95+110 (2020)
12. Liu, S., Liu, X., Wang, S., Khan, M.: Fuzzy-Aided solution for Out-of-View challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
13. Chang, Y., Li, W., Ba, Y., et al.: A new method for frequency control performance assessment on operation security. *Trans. China Electrotech. Soc.* **34**(06), 1218–1229 (2019)
14. Xu, D., Wang, D.: resonance frequency tracking control of magnetic resonance wireless power transfer systems. *Autom. Instrum.* **17**(04), 20–24 (2020)
15. Zhao, Z., Meng, Z., Zhang, J., et al.: Control strategy of coordination and equalization for diesel generator and energy storage system in emergency microgrid. *Autom. Electr. Power Syst.* **43**(10), 53–59+141 (2019)



# PID Based Automotive Electronically Controlled Power Steering System in the Internet of Things Environment

Qing-huan Qin<sup>(✉)</sup> and Hong-yu Huang

Guangxi Modern Polytechnic College, Hechi 547000, China  
liuzhengyuan7788@163.com

**Abstract.** It is difficult to adjust parameters automatically to obtain the optimal control parameters, and the steering performance of the system is often not ideal. In order to solve this problem, this paper designs an automotive electronic control type power steering system based on PID under the Internet of Things environment. A switching step-down circuit is composed of a combined power supply mode, and then the drive circuit is designed, and the output end is flexibly configured through the control pin. In the software design part, the vehicle system dynamics model is established by analyzing the steering resistance torque, and then the proportional relationship between the power assist torque and the motor current is determined. Finally, the PID control algorithm is designed to set the control parameters automatically. Finally, the main program of the control system is set up, and the bus is used as the connecting medium to control the driving torque of the vehicle. The experimental results show that the system has good steering performance, and the test results of steering Angle, yaw velocity, lateral acceleration and offset path are better than those of the existing system, which is beneficial to the safe driving of the vehicle.

**Keywords:** Internet of Things · PID · Automotive electronic control · Power steering

## 1 Introduction

With the rapid development of the automotive industry, people have higher and higher requirements for vehicle performance. Among them, the performance that attracts the most attention is the smoothness and handling stability of the vehicle. The electronically controlled power steering system of a vehicle is a key component that affects the ride comfort and handling stability of the vehicle. Therefore, the research and development of the vehicle steering system has become the key for major manufacturers to improve the performance of their vehicles. During the steering of the vehicle, the power steering system assists the driver in steering the vehicle by providing steering power.

The steering system of a vehicle is very important for handling stability. The automobile steering system has gone through three development stages of pure mechanical steering system, hydraulic power steering system, and electronically controlled power steering system. The steering force of the mechanical steering system is all provided by people, and it is mainly composed of three parts: steering mechanism,

steering gear and steering transmission mechanism. The mechanical steering system has a simple structure and strong reliability, but it is quite laborious to use, and stability, accuracy, and safety cannot be guaranteed. The hydraulic power steering system adds a hydraulic device to the mechanical steering system. Under normal circumstances, most of the force required for car steering is provided by the hydraulic assist device, and the driver only provides a very small part of the torque as a signal [1]. Hydraulic steering system has the advantages of flexible operation, light weight, and can relieve tire impact. Since its inception, it has been widely used in automobiles. But the hydraulic power steering system cannot solve the coexistence of low-speed steering portability and high-speed steering stability. In addition, the hydraulic oil pump keeps working while the car is running, which increases fuel consumption, which restricts its own development. As a new generation of automotive steering systems, electronically controlled power steering systems are applied to various types of automobiles at a tremendous growth rate every year. In the early years, because of cost issues, the electronically controlled power steering system became synonymous with high-end cars. In recent years, electronic technology has developed rapidly and the prices of electronic components have been continuously reduced, making it possible to greatly reduce the cost of electronically controlled power steering systems and mass production. Electronically controlled power steering systems are also occupying more shares in automotive steering systems. Compared with traditional steering systems, electronically controlled power steering systems have many advantages, making them more and more popular. Its advantages are: it can provide the best assistance under various working conditions, which improves the fuel economy of the car; the system is directly assisted by the motor, which can also provide assistance when the engine is stalled or malfunctions; it has a compact structure and few parts. It is easy to install and assemble; it has good low-temperature working performance, and has higher reliability and safety; it can greatly reduce development costs and improve development efficiency. It only depends on changing the control strategy of the system to meet the needs of different models and different drivers.

In the process of vehicle steering, in order to improve the steering portability and steering stability of the vehicle, reasonable and effective control of the electronically controlled power steering system is the key to ensuring its good performance [2]. PID control controls the torque output of the booster motor by converting the current signal into the control voltage signal, so as to achieve the control effect of the electronically controlled power steering system, and improve the smoothness of the vehicle, the steering stability, and the ease of steering. Comprehensive performance is of great significance. Therefore, this article designs a PID based automotive electronically controlled power steering system under the Internet of Things environment to promote the industrial development of my country's automotive power steering system and enhance the international competitiveness of the automotive industry. The design idea of this system is as follows:

- (1) The combined power supply mode is used to form a switching step-down circuit by using the internal switching mode and the external circuit. Then the drive circuit is designed and the output end is flexibly configured through the control pins.

- (2) The steering resistance moment was analyzed and the vehicle system dynamics model was established. After determining the proportional relationship between the power torque and the motor current, the power steering control algorithm was designed based on PID, so as to automatically set the control parameters.

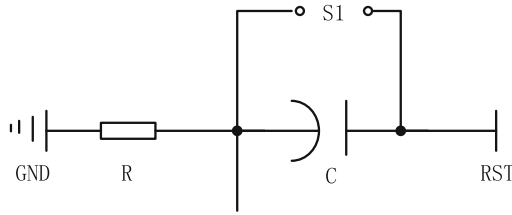
## **2 The Hardware Design of the Automotive Electronically Controlled Power Steering System Based on PID in the Internet of Things Environment**

The hardware circuit of the automobile electronically controlled power steering system is the basis for the implementation of advanced engine control methods and control strategies. Therefore, research and development of a reliable and good-performance hardware circuit board is of great significance. The main modules of the hardware circuit are analyzed below.

### **2.1 Controller Selection and Circuit Design**

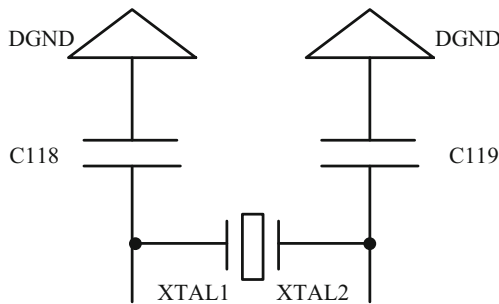
In order to simplify the design of the hardware circuit board of the automotive electronically controlled power steering system and improve the operational reliability, a microcontroller with rich on-chip resources should be selected. On-chip resources include A/D converters, communication interfaces CAN, SPI, voltage converters, etc. Based on the principle of being able to meet design requirements and not pursuing extreme performance, this article selects Infineon's microcontroller SAK-TC1782F-320F180HL as the main control chip. TC1782 is an excellent high-performance microcontroller, used in complex embedded control systems, mainly used in engine and gearbox control. It has the characteristics of high cost performance, strong computing power, and fast real-time response. The highest frequency of TC1728 can reach 180 MHz, and the internal integration of 2.5M PFlash and 128K Dflash. TC1782 uses reduced instruction set computing (RISC) processor architecture, which achieves high computing bandwidth at a lower system cost. It integrates digital signal processing (DSP) operation and addressing mode. This mode provides powerful computing power and can effectively analyze various complex signals in the real world [3]. TC1728 has 2 independent ADC modules, a total of 32 analog signal input channels with a conversion accuracy of 10 bits, and a total of 86 digital general-purpose I/O ports. TC1782 also integrates TriCore CPU, multiple buses, bus arbitration, interrupt controller, peripheral control processor, DMA controller and multiple on-chip peripherals. Whether the TC1782 microcontroller can work normally depends on the quality of the minimum system circuit design. The minimum system circuit is composed of power supply, reset, oscillation, external expansion circuit, and program download and debugging. The designed external key-press hardware trigger reset circuit is shown in Fig. 1.





**Fig. 1.** Reset circuit design

TC1782 is divided into warm reset and cold reset. The warm reset includes watchdog timer (WDT) trigger reset, debug (OCDS) trigger reset, software reset (SW) and external power-on hardware automatic reset, etc.; cold reset includes the above Electrical reset trigger request (PORST); TC1782 is a high-level reset chip. Pull the reset pin /PORST level within 0.1  $\mu$ S, and the microcontroller can reset once. VCC is the microcontroller power supply 5V, RST is connected to the reset pin of the microcontroller, the capacitance is a 10UF electrolytic capacitor, and the resistance is a 10K ceramic resistor. This circuit resets once when the control system is powered on, and the system resets again when the button is pressed [4]. Therefore, the operating system is reset by closing and opening the button. The crystal oscillator frequency supported by TC1782 is between 8–40 Hz, and the crystal oscillator circuit diagram designed is shown in Fig. 2.



**Fig. 2.** Crystal oscillator circuit design

The parameters of each component in the circuit are designed according to the values recommended by the manufacturer. The capacitance is 33 pF, the crystal oscillator is 20 MHz, and the two ends of the crystal oscillator are respectively connected to the XTAL1 and XTAL2 pins of the microcontroller. The circuit is connected to the oscillator inside the single-chip microcomputer through the above two pins to provide a square wave of the clock signal for the single-chip microcomputer, and the circuit has no active components, so the designed circuit is called a passive internal clock oscillation circuit.

## 2.2 Power Module Design

The power supply of the controller is an important guarantee for the normal operation of the power steering system, so the design of the power supply should meet the following requirements: (1) When designing the power supply voltage of the MCU I/O port, the compatibility with the input level of the peripheral drive circuit should be taken into account; (2) The voltage of the ADC module must be accurate, and the voltage value must be greater than or equal to the voltage value of the input sensor signal; (3) the current load capacity of the power supply circuit should meet the requirements; (4) In order to make the circuit low power consumption, stable performance and good electromagnetic compatibility, according to the actual situation to choose the best power supply mode [5].

The various modules of the controller require different power supply voltages. The CPU core requires 1.3 V to reduce power consumption, the I/O port module requires 3.3 V or 5 V power supply, and the ADC module requires 5 V power supply. Therefore, the controller requires a total of 5 V, 3.3 V and 1.3 V power supply. TC1782 integrates two voltage regulators, 3.3 V and 1.3 V. 1.3 V includes linear step-down mode and switching step-down mode. The external power supply module uses TLE7368-3E recommended by Infineon. This chip can output 5 V, There are 3 voltages of 3.3 V and 1.3 V, so the power supply mode of TC1782 is very flexible. Through the flexible combination of internal and external power supply modules, this article adopts the combined power supply mode recommended by Infineon. The external power supply is 5 V and 3.3 V, and the internal switching mode of the 1.3 V regulator and the external circuit form a switching step-down circuit. Because the difference between the external 5 V power supply voltage and 1.3 V is too large, if 1.3 V uses the linear step-down mode, the energy consumption inside the microcontroller is too large, causing the microcontroller to heat up, so the above recommended mode is used to power the microcontroller. TLE7368 is a power chip specially produced by Infineon for the Powertrain system. The chip can provide 5 V, 3.3 V and 1.3 V power supply voltages and functional safety testing for 32-bit microcontrollers. TLE7368 first uses a pre-regulator to output a 5.5 V pre-regulated voltage, and then uses a linear regulator to convert the voltage to 5 V, 3.3 V, and 1.3 V. This can reduce power consumption and improve voltage accuracy, and the output voltage accuracy is 2%. Among them, the current load capacity of 5 V and 3.3 V is 800 Ma. Most of the chips listed above are in switch regulation mode. Compared with linear regulation mode power supply, switch regulation mode has many advantages such as low power consumption, high efficiency, small size, light weight, and wide voltage regulation range.

## 2.3 Drive Circuit Design

The design of the hardware drive circuit module of the power steering system is compatible with the requirements of the vacuum diaphragm and linear circulation valve on the drive circuit. The dual H-bridge drive chip L9960 based on STMicroelectronics is used as the drive chip of the drive circuit. L9960 is a dual H type driver chip specially launched by STMicroelectronics for automotive-grade motor drive, mainly

used for power steering control. L9960 mainly has the following features: flexible configuration of the output terminal through the control pin; battery voltage range from 4.5 V to 28 V; chip power supply voltage from 4.5 V to 5.5 V; input switching frequency up to 20 kHz; input logic level compatible with 3.3 V and 5 V; integrated chip power supply monitoring module; integrated short-circuit protection, over-temperature protection module; integrated SPI interface and self-diagnosis module. The drive circuit module circuit designed based on the L9960 drive chip is simple, uses fewer components, works stably, and has better performance. The driving circuit is shown in Fig. 3.

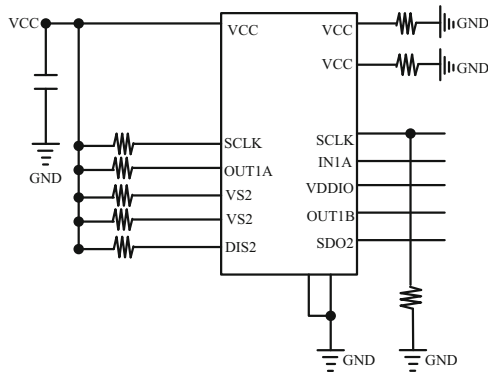


Fig. 3. Drive circuit design

When the actuator adopts the vacuum diaphragm type, a low-end output is used to control the current of the vacuum solenoid valve; when the linear valve is used as the actuator, an H bridge is used as the control of the solenoid coil. The working process of the vacuum diaphragm valve is: a valve stem and a spring are connected to both ends of the rubber membrane [6]. Under normal circumstances, the spring force makes the valve stem tightly against the valve seat, so that the valve is in a closed state. When the vacuum force on the rubber diaphragm in the direction of the vacuum port is greater than the spring force, the valve stem is pulled away from the valve seat, the EGR valve opens, and the exhaust gas enters the intake pipe from the exhaust pipe through the valve port. When the vacuum force on the rubber diaphragm in the direction of the vacuum port is less than the spring force, the rubber diaphragm drives the valve stem to fall back to the valve seat, and the valve port is closed. Therefore, the opening of the valve port is controlled by the degree of vacuum, and the degree of vacuum is controlled according to the actual working conditions of the engine, so it is also indirectly controlled by the vacuum solenoid valve. There is a solenoid coil group and an armature inside the linear valve. The electromagnetic coil is driven to generate electromagnetic force, so that the armature shaft pulls the cone valve up, the valve port opens, and the exhaust gas flows from the exhaust manifold into the intake manifold. The valve port position sensor is actually a linear potentiometer, and its range is between 1–4.5 V from closed to open. The optimal EGR rate is calculated according to

various parameters of the engine, combined with the position sensor feedback position signal, and the on-off of the solenoid current is controlled by outputting a PWM modulation signal to achieve linear and precise control purposes. Compared with the vacuum diaphragm valve, although the linear valve has a complicated structure and a higher cost, it has unique advantages in control accuracy and closed-loop control, so it is also widely used in many high-end cars. According to the control requirements of the power steering system, a microcontroller is selected and a reasonable control circuit is designed.

### 3 PID Based Automotive Electronic Control Power Steering System Software Design in the Internet of Things Environment

#### 3.1 Establish Vehicle System Dynamics Model

The automobile electronically controlled power steering system is actually a complex system composed of many non-linear links. In order to ensure the correctness of the established model, it is necessary to simplify the secondary factors and uncertain conditions in the mathematical model. This chapter establishes a dynamic model for the power steering system. The force in the power steering system is mainly the steering force acting on the steering wheel by the driver, the assist torque of the assist motor, and the internal and external resistance torque of the entire steering system. The driving force of the driver mainly includes the steering force when steering the car and the holding force during turning. The resistance torque of the steering system mainly includes the internal resistance torque of the steering system and the external resistance torque generated by the tires. Among them, the internal resistance moment of the steering system includes the friction resistance moment of the steering system, the restoring moment of the steering system, the inertia moment of the steering system, etc.; the external resistance moment is the resistance moment generated by the tire and the ground around the kingpin [7]. In order to obtain accurate dynamic characteristics of system steering and manipulation, a mathematical model of the system must be built, that is, a mathematical expression of the dynamic characteristics of the system. The correct system mathematical model can also be used for the selection and calculation of system component parameters, guide the software and hardware development of the electronic control unit, and can also be used for subsequent research on the influence of control parameters on vehicle handling and stability, and for the construction of subsequent test benches. All have important guiding significance. According to the structure of the power steering system, the power steering system can be divided into four components: steering wheel and steering input shaft, steering output shaft, rack and motor. The assist torque of the electric motor acts on the steering output shaft. According to Newton's law, the established dynamic equation of steering system is:

$$w\alpha + e\beta = s + hj - kr \quad (1)$$

In formula (1),  $w$  represents the stiffness of the sensor torsion bar;  $\alpha$  represents the angle between the steering wheel and the steering input shaft;  $e$  represents the damping coefficient of the steering shaft;  $\beta$  represents the angle between the steering wheel and the steering output shaft;  $s$  represents the detected torque;  $h$  represents the reduction ratio of the reducer;  $j$  represents the electromagnetic torque of the motor;  $k$  represents the equivalent force of the tire return torque on the rack;  $r$  represents the radius of the pinion. The torque sensor is based on the angle difference between the steering wheel angle and the lower steering output shaft as the measurement object, so it can be simplified as a torsion spring, so the calculation formula for the detected torque is:

$$s = w(\theta_1 - \theta_2) \quad (2)$$

In formula (2),  $\theta_1$  represents the steering shaft angle;  $\theta_2$  represents the steering pinion angle. In this way, the normalizing torque and the cornering angle of the front and rear wheels can be changed in time with the changes in the driving state of the vehicle, and the force is more in line with the actual situation. To get the steering wheel torque of a car, you must first know the tire's aligning torque, and the tyre's aligning torque is related to the tire slip angle. The tire slip angle is the car's center of mass slip angle, yaw rate, body roll angle, and A function of the front wheel angle [8]. Therefore, a complete vehicle model of motor sports must be established. The movement of the car can be described by the Cartesian coordinate system of OXYZ vehicle coordinate system. XOZ is in the symmetrical plane of the car. When the vehicle is at a standstill on a level road, the X axis is parallel to the ground and points forward, the Z axis points upward through the center of mass, and the Y axis points to the driver's left side. The origin O of the coordinate system coincides with the center of mass. The entire coordinate system conforms to the right-handed spiral rule. When the car is turning, the body roll will cause the vertical load of the inner and outer tires to change, which affects the lateral force and the torque of the tire. In this chapter, through the analysis of the steering resistance torque of the power steering system, the relationship expression between the steering wheel torque and the steering resistance torque is established, which lays the foundation for the research of the control strategy and the construction of the simulation model.

### 3.2 Set the Parameters of the Assist Characteristic Curve

The assist characteristic refers to the relationship between the steering wheel torque and the assist torque or the assist current. After the power steering system receives the steering wheel torque signal and the vehicle speed signal, it can determine the value of the assist current applied to the motor according to the assist characteristics. For a booster motor, the actual boost torque of the motor is controlled according to the delivered boost current. For the electric power steering system, the magnitude of the assist torque provided by it is proportional to the current of the motor, so the assist characteristic is represented by the relationship between the motor current, vehicle speed and steering wheel torque [9]. The ideal assist characteristics should be able to fully coordinate the steering ease of the vehicle and the perception of the road surface. Under the premise of satisfying the steering portability of the vehicle, the road feel does

not change much, so that the driver can judge the change of the vehicle driving condition. In order to make the driver handy in the steering process of the vehicle, the electric power steering system should try to meet the driver's original driving habits. The basic requirement that the assist characteristics should meet is the relationship between the steering wheel torque and the assist torque. Under normal circumstances, there are three main forms of the assist characteristic curve, linear type, polyline type and curved type. The linear power assist characteristic curve is the most intuitive, which can make the design of the control system easier and easier in the control process. However, because the output of the assist characteristic curve is linear, the driver feels uniform on the road surface. The curve-shaped power assist characteristic curve is more complicated, and a large amount of data is needed for support in the process of determining the curve, and it is not easy to adjust in the actual use process, so this method is not used in this article. The characteristic of the broken line assist characteristic curve lies between a straight line and a curved line. In summary, the design of the linear boost characteristic curve is simple and easy to control, so this paper selects the linear boost characteristic curve. In the broken-line assist characteristic curve, in the part of the assist change, the steering assist current has a piecewise linear relationship with the steering wheel torque, and its function expression is as follows:

$$C = \begin{cases} 0, & 0 \leq J < J_0 \\ a(J - J_0), & J_0 \leq J < J_1 \\ a_2(J - J_1) + a_1(J_1 - J_0), & J_1 \leq J < J_{\max} \\ C_{\max}, & J \geq J_{\max} \end{cases} \quad (3)$$

In formula (3),  $C$  represents the assist current of the electric motor;  $a, a_1, a_2$  represents the gradient of the assist characteristic curve;  $J$  represents the torque input by the steering wheel;  $J_0$  represents the input torque of the steering wheel when the power steering system starts assisting;  $J_1$  represents the assist characteristic The steering wheel input torque when the curve gradient changes from  $a_1$  to  $a_2$ ;  $J_{\max}$  represents the steering wheel input torque when the power steering system can provide the maximum assist;  $C_{\max}$  represents the maximum operating current of the electric motor. According to the established assist characteristic curve, three basic characteristic parameters need to be determined, namely the input torque  $J_0$  of the steering wheel when the system starts assisting, the gradient coefficient  $a$  of the assist characteristic curve and the steering wheel input rotation when the maximum assist is provided. When the vehicle is turning, the driver turns the steering wheel to make the vehicle make a steering movement. In this process, the force acting on the steering wheel should be within an appropriate range, neither too large nor too small. According to the sensitivity of human perception, when the steering torque input by the steering wheel is less than a specific value  $J_0$ , then the assist motor does not provide steering assistance. This method is also helpful to reduce fuel consumption, so the value of  $J_0$  is 1. The assist torque output by the assist motor has a maximum upper limit. When the steering wheel input torque reaches a certain value, the maximum torque is output when the control current reaches the maximum. At this time, the steering assist torque will not continue to increase, so that the motor can be Some protection, so the value of  $J_{\max}$  is 7. At different vehicle speeds, the value of the gradient coefficient  $a$  of the assist

characteristic curve is different, and the value decreases as the vehicle speed increases. This value indicates that the driver can fully feel the road surface information while ensuring the ease of steering while the vehicle is driving. The gradient coefficient of the assist characteristic curve set in this paper is shown in Table 1.

**Table 1.** Gradient coefficient table of boost characteristic curve

Speed	Gradient coefficient	Speed	Gradient coefficient
0	3.352	40	1.413
10	2.278	50	1.269
20	1.964	60	1.118
30	1.657	70	0.926

According to the selection of the parameter values of the above-mentioned assist characteristic curves, the design of the assist characteristic curve is completed, and a linear relationship between the assist current of the assist change part and the steering wheel torque is obtained.

### 3.3 Design of Power-Assisted Steering Control Algorithm Based on PID

The specific control process of the automobile electronically controlled power steering system is: the vehicle turns during driving, and the steering angle signal input by the steering wheel is converted into a torque signal. The power steering system receives the torque signal transmitted by the steering wheel and the vehicle speed signal and processes and analyzes it to output a current signal. This signal is the input signal of the PID controller by subtracting a current signal generated by the power-assisted motor during operation. The PID controller converts this signal into a control voltage signal through proportional-integral-derivative control to control the torque output of the booster motor, and then achieves the control effect of the electric power steering system through the actuator. In the process of using the PID controller to control, the proportional control link proportionally reflects the deviation signal of the control system. Once the deviation occurs, the controller immediately takes control to reduce the deviation, play the main control role, and make the system stable; The integral link is mainly used to eliminate the static error and improve the error-free degree of the system. The strength of the integral action depends on the integral coefficient. The smaller the integral coefficient, the weaker the integral action, and vice versa. The differential link reflects the change trend of the deviation signal, and can introduce a system into the system before the deviation signal becomes too large. Correct the signal early to speed up the action speed of the system, adjust quickly to eliminate the deviation and reduce the adjustment time. When the three control parameters are adjusted properly, the advantages of PID control can be fully utilized to obtain the ideal control effect. Its expression is:

$$Y = \lambda_1 + \frac{\lambda_2}{s} + \lambda_3 s \quad (4)$$

In formula (4),  $Y$  represents the transfer function;  $\lambda_1$  represents the proportional coefficient;  $\lambda_2$  represents the integral coefficient;  $\lambda_3$  represents the differential coefficient;  $s$  represents the deviation. In order to complete PID control, three parameter values need to be set. Conventional parameter tuning methods are mainly proposed based on the first-order inertia model with delay. In practical applications, the tuning of the three parameters can be obtained by “trial and error” on the simulation model. The order of trial and error is carried out in accordance with the principle of first proportional, then integral, and then differential. However, in actual operation, the PID parameters are obtained by trial and error in a purely proportional manner. The trial and error process is highly repeatable, low in efficiency and has a lot to do with the experience of the technicians [10]. And in fact, there are many uncertain factors in the power steering system, such as random impact on the road surface and friction within the system. The parameters of the traditional PID control after debugging can only achieve good results in this state, and it is difficult to automatically adjust the parameters to obtain the optimal control parameters in the whole process, and the system control effect is often not ideal. Aiming at this problem, this paper uses sliding mode variable structure to improve the PID control algorithm [11]. The sliding mode deformation structure can force the system to move up and down with a small amplitude and high frequency along a prescribed state track under certain characteristics, which has nothing to do with the parameters and disturbances of the system, so the system can have good robustness. The approach speed of the exponential approaching law can be changed, and it has the advantages of not only accelerating approach time but also reducing chattering. Therefore, this article adopts the exponential approaching law control method, and its expression is:

$$l = -\eta \text{sgn}(v) - bv \quad (5)$$

In formula (5),  $l$  represents the output of the controller;  $\eta, b$  represents the reaching law coefficient;  $v$  represents the switching function, which is the error between the reference current and the actual current. In this case, the progressiveness of the sliding mode movement is stable and the dynamic quality is good [12]. Through the optimization of PID algorithm, the problem of PID control difficult to self-tuning control parameters can be better solved.

### 3.4 Set up the Control System Program

The design of the main program is the key and basic work for the establishment of system software. The initialization of the system is mainly the initialization of the port, the initialization of the area, and many other aspects. When the main program starts to operate, it will use multiple sensors of the system, such as acceleration and brake pedal sensors, clutch switch sensors, and gear switch input signals to enter the judgment mode and start the corresponding work. Then start each program mode and perform related operations. During the operation, the main program runs in the form of a loop,



and the change of the vehicle operation mode will cause the system to be interrupted. Use the controller to process the signals in the 0 and 5V voltage range from the acceleration and brake pedal position sensors, and use the A/D converter as the platform to realize the signal conversion. After the conversion operation is successfully completed, the corresponding signal will be obtained [13]. This result is mainly used to clarify and judge the driving condition of the car.

When the switch is closed, the system uses the engine as the main power of the vehicle and converts to the engine drive mode [14]. The difference between the engine speed and the starting speed should be defined at the beginning. If it is greater than the latter speed, then the system will not have a start command action. If the engine speed is less than the value of the starting speed, the system will judge the pressure of the accumulator at this moment. If the pressure of the accumulator is greater than the critical value, the hydraulic power system will be used to start the engine. If it is less than the critical value, the starter generator will be used to complete the start of the generator [15]. When the car's speed drops to zero, the engine will no longer continue to work at this moment, the car's gear is in a non-neutral state, the clutch angle is reduced, the accelerator pedal is gradually released, and the system will start the single hydraulic drive mode.

The main purpose of the electronic control unit is to complete the detection of the pressure of the accumulator. If the pressure signal is greater than the pre-required pressure threshold value, the detection will be initiated, and the detection of the displacement signal of the accelerator pedal will be performed at the same time. According to the test results, the variable displacement hydraulic pump and motor are calculated. When the speed is not zero, the gear is not neutral, the clutch is combined, the accelerator pedal is pressed, and the system will start the commonly used engine drive mode at this moment. With the bus as the connection medium, the power steering system will be connected to the engine, and the engine will perform the task of controlling the driving torque of the car at this moment. At this point, the design of each module of the software is completed. Combined with the hardware part, complete the design of the automotive electronically controlled power steering system.

## 4 Experimental Study

This paper designs a PID based automotive electronically controlled power steering system under the Internet of Things environment. In order to test the performance of the system, the following experiments are carried out to determine the effectiveness of the power steering system.

### 4.1 Experiment Preparation

According to the structure and performance parameters of the key components of the automobile studied in this paper, the parameters required in the simulation system are set, as shown in Table 2.

**Table 2.** Simulation parameter table

Serial number	Parameter name	Numerical value
1	Moment of inertia of steering wheel ( $\text{kg}\cdot\text{m}^2$ )	0.015
2	Steering column damping ( $\text{N}\cdot\text{m}\cdot\text{s}/\text{r}$ )	0.030
3	Equivalent moment of inertia of steering mechanism and front wheel ( $\text{kg}\cdot\text{m}^2$ )	0.015
4	Equivalent damping of steering mechanism and front wheel ( $\text{N}\cdot\text{m}\cdot\text{s}/\text{r}$ )	0.05
5	Steering gear ratio	25
6	Steering gear positive efficiency (%)	85
7	Torque sensor equivalent torsion bar stiffness ( $\text{N}\cdot\text{m}/\text{deg}$ )	2.5
8	Reducer transmission ratio	20
9	Reducer efficiency (%)	85
10	Moment of inertia of booster motor ( $\text{kg}\cdot\text{m}^2$ )	0.01
11	Assist motor damping ( $\text{N}\cdot\text{m}\cdot\text{s}/\text{r}$ )	0.03
12	Electromagnetic torque constant of booster motor ( $\text{N}\cdot\text{m}/\text{A}$ )	0.07
13	Assisted motor back electromotive force constant ( $\text{V}\cdot\text{s}/\text{r}$ )	0.015
14	Power-assisted motor resistance ( $\Omega$ )	0.2
15	Booster motor inductance (H)	0.001

Under the condition of co-simulation, the steering angle step input is carried out, and the PID control parameters are adjusted. The determined PID control parameters are as follows: proportional coefficient 0.6, integral coefficient 0.0045 and differential coefficient 0.0001. After setting, the output current of the PID controller has a good consistency in tracking the target current. Under the above experimental platform and parameter settings, experiments are carried out to test the performance of the auto-motive electronically controlled power steering system.

## 4.2 Experimental Results and Analysis

In order to verify the application effect of this design system, it is compared with the existing automobile power steering system. This article mainly evaluates the steering performance of each system. The steering performance describes the various dynamic responses of the car under the driver's steering command, as well as the corresponding steering power, steering wheel force feedback, and steering feel. The steering performance can be divided into four evaluation areas: steering characteristics when driving in a straight line, steering characteristics when driving in a curve, steering wheel force characteristics, and response shock caused by steering. Through the analysis of steering performance, four parameters of steering wheel angle, yaw rate, lateral acceleration and offset path are selected as indicators of steering performance to evaluate the steering performance of each system. The experimental test results are shown in Tables 3, 4, 5 and 6.

**Table 3.** Test results of steering wheel angle (N·m)

Serial number	Design system	Existing system 1	Existing system 2
1	20.52	38.42	39.21
2	21.24	38.26	39.64
3	20.65	38.58	39.47
4	20.88	38.47	39.54
5	20.76	38.74	39.76
6	21.17	38.49	39.65
7	21.22	38.55	39.87
8	20.58	38.62	40.28
9	20.63	38.54	40.14
10	20.84	38.33	39.88

According to the steering wheel angle test results in Table 3, the average steering wheel angle of the design system in this paper is 20.85 N·m, and the average steering wheel angle of the existing design system is 38.50 N·m and 39.74 N·m, indicating that the system is designed in this paper during the steering process. The steering torque is smaller, and the power-assisted control effect is better and obvious.

**Table 4.** Test results of yaw rate (deg/s)

Serial number	Design system	Existing system 1	Existing system 2
1	6.88	8.26	9.64
2	6.56	8.37	9.88
3	6.65	8.55	9.85
4	6.62	8.42	9.77
5	6.84	8.34	9.65
6	6.91	8.62	9.56
7	6.75	8.58	9.88
8	6.87	8.42	9.64
9	6.54	8.53	9.71
10	6.88	8.68	9.62

According to the yaw rate test results in Table 4, the average yaw rate of the system designed in this paper is 6.75 °/s, and the average yaw rate of the existing design system is 8.48 °/s and 9.72 °/s. The yaw rate of the system designed in this paper is 8.48 °/s and 9.72 °/s. The angular velocity is less than that of the existing system, which shows that the system designed in this paper can improve the handling stability.

**Table 5.** Lateral acceleration test results (g)

Serial number	Design system	Existing system 1	Existing system 2
1	1.36	2.62	3.23
2	1.25	2.84	3.35
3	1.37	2.87	3.46
4	1.32	2.78	3.33
5	1.58	2.66	3.54
6	1.34	2.85	3.48
7	1.59	2.64	3.62
8	1.24	2.51	3.59
9	1.23	2.64	3.42
10	1.45	2.83	3.35

According to the results of the lateral acceleration test in Table 5, the average lateral acceleration of the system designed in this paper is 1.37 g, and the average lateral acceleration of the existing design system is 2.72 g and 3.44 g, indicating that the system designed in this paper has good power-assisted characteristics and is easy to steer. Sex is strengthened.

**Table 6.** Offset path test results (m)

Serial number	Design system	Existing system 1	Existing system 2
1	2.62	3.82	4.23
2	2.54	4.03	4.34
3	2.75	4.14	4.45
4	2.67	3.97	4.58
5	2.88	3.95	4.39
6	2.56	3.86	4.42
7	2.44	3.68	4.35
8	2.55	3.77	4.27
9	2.68	3.95	4.26
10	2.82	3.82	4.34

According to the test results of the offset path in Table 6, the average offset path of the designed system in this paper is 2.65 m, and the average offset path of the existing design system is 3.90 m and 4.36 m. The path offset of the designed system in this paper has a significant reduction. Small, indicating good obstacle avoidance ability during driving.

Based on the above experimental test results, the PID based automotive electronically controlled power steering system designed in this paper has good steering performance. The test results of the four parameters of steering wheel angle, yaw rate, lateral acceleration and offset path are better than the existing ones. The system can

improve the handling stability and obstacle avoidance ability. The power-assisted steering effect of the designed system is better, which is conducive to the safe driving of the car.

## 5 Conclusion

In this paper, the design of automotive electronic control type power steering system based on PID under the Internet of Things environment is carried out from the hardware and software aspects. In the design process, the combined power mode is firstly adopted, and the internal switching mode and the external circuit are used to form a switching step-down circuit. Then the drive circuit is designed, and the output end is flexibly configured through the control pin. Based on this, the steering resistance torque is analyzed and the vehicle system dynamics model is established. After determining the proportional relationship between the power torque and the motor current, the power steering control algorithm is designed based on PID, and the control parameters are automatically set. The experimental results show that the system has good steering performance and good power steering effect, which is beneficial to the safe driving of the vehicle.

However, the work in this paper is only a theoretical simulation study, it is a pity that the designed system has not been tested on the bench and the actual vehicle, the effectiveness of the model and the accuracy of the control need to be tested in practice. These need to be solved in the future research.

**Fund Projects.** 1. 2020 Guangxi University Young and middle-aged teachers' scientific research basic ability promotion project: Research on classroom teaching dynamic evaluation system based on mobile terminal (2020ky45011).

2. 2016 Guangxi Vocational Education Teaching Reform Project: Study on the Connection of Project Design between Specialized Courses in "Project Teaching" of Applied Electronic Technology Major in Higher Vocational Education, (GXGZJG2016B003).

## References

1. Li, S., Guo, L., Cui, G., et al.: Return control of electronic power steering unequipped with an angle sensor. *Int. J. Automot. Technol.* **20**(2), 349–357 (2019)
2. Yamamoto, K., Sename, O., Koenig, D., et al.: Design and experimentation of an LPV extended state feedback control on Electric Power Steering systems. *Control. Eng. Pract.* **90** (9), 123–132 (2019)
3. Na, S., Li, Z., Qiu, F., et al.: Torque control of electric power steering systems based on improved active disturbance rejection control. *Math. Probl. Eng.* **2020**(6), 1–13 (2020)
4. Murilo, A., Rodrigues, R., Teixeira, E.S., et al.: Design of a parameterized model predictive control for electric power assisted steering. *Control Eng. Pract.* **90**, 331–341 (2019)
5. Chen, J., Guo, Y.: Research on control strategy of the electric power steering system for all-terrain vehicles based on model predictive current control. *Math. Probl. Eng.* **2021**(3941), 1–15 (2021)

6. Bai, H., Liu, C., Ma, R., Paire, D., Gao, F.: Device-level modelling and FPGA-based real-time simulation of the power electronic system in fuel cell electric vehicle. *IET Power Electron.* **12**(13), 3479–3487 (2019)
7. Petros, K., Geyer, T., et al.: Guidelines for the design of finite control set model predictive controllers. *IEEE Trans. Power Electron.* **35**(7), 7434–7450 (2019)
8. Li, Y., Feng, Q., Zhang, Y., et al.: Dual-motor full-weight electric power steering system for commercial vehicle. *Int. J. Automot. Technol.* **20**(3), 477–486 (2019)
9. Chung, S.S., Lee, Y.Z., Kim, E.S.: Development of the wear resistant ball nut assembly in the rack-based electric power steering. *Int. J. Automot. Technol.* **21**(3), 713–718 (2020)
10. Li, W., Chen, J., Lan, F., et al.: Exploitation of the electromechanical coupling system for automobile electric power-assisted steering. *J. Mach. Design* **37**(5), 8–14 (2020)
11. Liu, S., Liu, X., Wang, S., et al.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
12. Liu, S.: Introduction of key problems in long-distance learning and training. *Mobile Netw. Appl.* **24**(1), 1–4 (2019)
13. Liu, S., Liu, D., Muhammad, K., et al.: Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* **16**(1), 66–75 (2019)
14. Xv, F., Liu, X., Chen, W., et al.: Fractional order PID control for steer-by-wire system of emergency rescue vehicle based on genetic algorithm. *J. Central South Univ.* **26**(9), 2340–2353 (2019)
15. Zhang, Z., Tang, L., Hao, W., et al.: Differential power steering control for in-wheel motored electric vehicle based on variable universe fuzzy PID. *J. Autom. Safety Energy* **10**(2), 169–177 (2019)



# Research on Debounce Method of Electronic Imaging Equipment Based on Feature Point Matching

Xiao-jing Qi<sup>(✉)</sup>

Chongqing Telecommunication Polytechnic College, Chongqing 402247, China  
qixiaojing2021@163.com

**Abstract.** Aiming at the problem of image generation blur in existing electronic imaging equipment de-shake methods, this paper proposes a method of electronic imaging equipment de-shake based on feature point matching. First, perform feature point matching on the motion path of the electronic imaging device, obtain the motion characteristics of the device using constraint conditions, and implement the motion path feature matching processing of the electronic imaging device in combination with the matching feature window. Secondly, use the two-dimensional spatial relationship of the image to perform rotation, translation, and zoom processing on the resulting image. Based on the electronic imaging equipment operating model, the six-parameter radiation method is used for motion estimation, and the motion curve characteristics are determined and modified according to the base function synthesis curve. The optimized processing of the estimated value then obtains the de-shake result of the electronic imaging device. The experimental results show that the PSNR value is about 72.64 dB and the MSE value is about 12.54 when the method in this paper is used for de-jitter processing, both of which are better than the traditional method. It can be seen that this method has a better de-jitter effect.

**Keywords:** Feature point matching · Electronic imaging equipment · Debounce · Image feature points

## 1 Introduction

With the increasing demand for images in various fields, many scholars and businesses have conducted a lot of effective exploration and research on electronic imaging equipment. With the increasing application of electronic imaging equipment, the equipment becomes more and more diversified, which can adapt to any scene, especially in a dynamic environment, it is easier to cause the electronic imaging equipment to shake and cause the image to blur. For this, de-shake It has become an urgent problem in this field [1].

Literature [2] proposes a de-jitter method based on a self-reference source, which does not require two similar sources to be measured under the premise of satisfying high measurement accuracy, which greatly reduces the complexity of the measurement system. Then, the structure of the measurement system is optimized, the mechanism of the artifact peak in the optical fiber delay line system is explained, and the second-order

difference frequency comb model is proposed to realize the de-jitter processing. Experimental results show that this method can effectively improve the accuracy of de-shake, but ignores the time-consuming situation; Literature [3] proposes to use the de-shake blur algorithm to restore the original image information of the blurred image, and then perform point-based restoration based on the motion recovery structure. The dense three-dimensional reconstruction of the cloud, and finally the Poisson surface reconstruction of the densely reconstructed point cloud to obtain the surface density and realize the de-jitter processing. The experimental results show that the de-jitter effect of this method is obvious, but it takes a long time. In this regard, this paper proposes a research on the de-shake method of electronic imaging equipment based on feature point matching. This method can not only solve the problem of blurred image, but also shorten the time of de dithering and improve the efficiency of de dithering method. The practical effect is good.

## 2 Feature Point Matching

In the process of electronic imaging equipment de-shake, this paper uses feature point matching algorithm for de-shake processing, but in fact, in the process of feature matching, the movement path of electronic imaging equipment directly affects the matching of feature points. The motion vector of the electronic imaging device is calculated [4]. Because the mapping transformation is transitive, the motion path of the electronic imaging device in the  $t$ -th frame can be represented by  $C(t)$ :

$$C(t) = F(t)F(t-1)\dots F(0) \quad (1)$$

In order to match the estimated operating path  $C(t)$  of the electronic imaging device with the collected feature points, it is necessary to reduce the jitter caused by the movement of the electronic imaging device as much as possible, and this needs to be restricted:

$$O\{P(t)\} = \sum_t \|P(t) - C(t)\|^2 + \sum_t \left( \lambda_t \sum_{r \in \Omega_t} \omega_{t,r} \|P(t) - P(r)\|^2 \right) \quad (2)$$

where:  $\Omega_t$  is the time-domain smoothing radius,  $\omega_{t,r}$  is the Gaussian weight, and  $\lambda_t$  is the balance factor. Using the motion path algorithm of a single device, the motion characteristics of the device in each grid are estimated:

$$C_i(t) = f_i(t) + f_i(t-1) + \dots + f_i(0) \quad (3)$$

where:  $f_i(t)$  represents the motion vector of the  $i$ -th grid.

The search starts from the center of the search area, that is, the origin. The search path uses a diamond search path to calculate the SAD value of each point [5]. The calculation formula is as follows:



$$SAD(i,j) = \sum_{m=1}^M \sum_{n=1}^N |I_1(m,n) - I_2(m+i+p,n+j+p)| \quad i,j = -p, \dots, p \quad (4)$$

Find the minimum SAD value, then its corresponding position is the matching feature window, and its center point is the matching feature point. The PROSAC algorithm is used to match the characteristics of the motion path, and the formula is as follows:

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \quad (5)$$

### 3 Research on Debounce Method of Electronic Imaging Equipment

It is assumed that the motion form of any electronic imaging device can be represented by rotation, translation, and zoom. Among them, the zoom movement of the electronic imaging device includes the changes in the depth of field caused by the forward and backward movement of the carrier and the imaging changes caused by the changes in the parameters of the electronic imaging device [6]. Therefore, the movement of the electronic imaging device can be described by the following formula, let  $(X, Y, Z)^T$  be a point on the scene in the coordinate system of the electronic imaging device,  $(X, Y, Z)^T$  and  $(X', Y', Z')^T$  are the three-dimensional coordinates of the same point at different times, and the corresponding point coordinates on the imaging plane are  $(x, y)$  and  $(x', y')$ . If the movement of the electronic imaging device is displacement, rotation and linear change, the relationship between  $(X, Y, Z)^T$  and  $(X', Y', Z')^T$  is:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \quad (6)$$

The translational movement of the electronic imaging device means that the device only has displacement in the  $x$  and  $y$  directions in two-dimensional space. If the device only has a translational movement, a translation model can be used:

$$p' = W(p, T) = p + T = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (7)$$

In the formula:  $p = (x, y)^T$  and  $p' = (x', y')^T$  respectively represent the coordinates of the equipment operating point at the reference point and the current  $K$ -th point, and  $t_x$  and  $t_y$  are the offsets of  $p'$  relative to  $p$  on the  $x$ -axis and  $y$ -axis, respectively.

In addition to the translational movement, the electronic imaging device itself may also undergo rotational movement. In this regard, the zooming movement is described by the zoom factor  $s$ . Therefore, when the electronic imaging device may have translation, rotation, or zooming motion, the Similarity motion model can be used:

$$P' = s \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} P + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \tag{8}$$

In the formula:  $\theta$  is the rotation angle and  $s$  is the zoom factor.

Connect the  $k$ -th point coordinate with the curve of the electronic imaging coordinate point, and determine the row and column displacement vector value of the current electronic device relative to the reference device according to the trough values of the two curves [7], the calculation formula is:

$$C(w) = \sum_{j=1}^N [Col_k(j+w) - Col_r(j+M)]^2 \quad 1 < w \leq 2M+1 \tag{9}$$

where:  $Col_k(j)$  and  $Col_r(j)$  are the projection values of the  $k$ -th point coordinates and the  $j$ -th column of the reference coordinate respectively,  $N$  is the length of the column, and  $M$  is the search width of the displacement vector relative to the reference coordinate point on one side. Assuming that  $W_{min}$  is the value of  $w$  when  $C(w)$  is minimum, the displacement vector of the  $k$ -th point coordinate relative to the reference coordinate in the vertical direction is:

$$\delta_c = M + 1 - W_{min} \tag{10}$$

Assume that  $f_1$  and  $f_2$  are two coordinate points that differ only by  $(d_x, d_y)$  in position, and  $f_2(x, y) = f_1(x - d_x, y - d_y)$ . The corresponding Fourier functions are  $F_1(w_x, w_y)$  and  $F_2(w_x, w_y)$ . According to the properties of the Fourier function, there are:

$$F_2(w_x, w_y) = e^{-j2\pi(w_x, w_y)} F_1(w_x, w_y) \tag{11}$$

Based on the phase shift characteristics of the Fourier function, suppose the cross power spectrum of the two coordinate points  $f_1$  and  $f_2$  is:

$$\frac{F_1(w_x, w_y) F_2^*(w_x, w_y)}{|F_1(w_x, w_y) F_2^*(w_x, w_y)|} = e^{-j2\pi(w_x, w_y)} \tag{12}$$

where:  $F_2^*$  is the conjugate function of  $F_2$ .

According to the constitution of the category variance, when  $\sigma_B^2 / \sigma_T^2$  is the largest, the de-jitter performance is the best, and the threshold at this time is the best threshold, and usually the threshold with the largest  $\sigma_B^2$  can be selected. Therefore, we select the

threshold  $T_r$ , that maximizes the variance between classes as the threshold of de-jitter  $R$ , namely:

$$T_r = \max_{T \in R} \left\{ P_{out} \sigma_{in} (\mu_{in} - \mu_{out})^2 \right\} \tag{13}$$

if the residual value of a point is less than  $T$ , it is a point within the coordinate, otherwise it is a point outside the coordinate. After removing all coordinate outer points, you can construct new coordinate points with the remaining inner points, and use this new point set to re-estimate the global debounce parameters [8]. Since the new point set is more in line with the law of global motion than the previous filtering, the re-estimated parameters will be more accurate.

In the construction of the operating model of electronic imaging equipment, a six-parameter radiation model is used for motion estimation [9]. Assuming that the initial value of the initial  $a_3, a_6, a_1, a_5$  of the translational movement of the electronic imaging device is 1, and the initial value of  $a_2, a_4$  is 0, the initial value of the intermediate parameter  $a_1^k \sim a_6^k$  can be calculated by the formula. Then use the iterative algorithm to obtain the final parameter values based on this starting point, so that the  $i$ -th point is compensated with the motion vector corresponding to this parameter to best match the background of the  $i - 1$ th point, that is, the minimum function value of the corresponding position is:

$$e = \frac{1}{n} \sum_{m=1}^n e_m^2, e_m = F_i(x_m, y_m) - F_{i-1}(x'_m, y'_m) \tag{14}$$

Among them,  $F_i(x_m, y_m)$  is the  $m$ -th feature of the coordinates of the  $i$ -th point, and  $F_{i-1}(x'_m, y'_m)$  is the  $m$ -th feature of the coordinates of the  $i - 1$ th point.

The given original  $N$  horizontal motion vectors  $u_i$  are independent one-dimensional curves, and the coordinates of each point on the curve are  $(x_i, u_i)$ , where  $(i = 0, 1, 2, \dots, N)$  uses the function  $p(x)$  to perform curve fitting, and the goal is to minimize the sum of squares of the error  $r_i = p(x_i) - u_i$  which is:

$$\sum_{i=0}^N r_r^2 = \sum_{i=0}^N [p(x_i) - u_i]^2 \tag{15}$$

In a geometric sense, it is to find the curve  $y = p(x)$  with the smallest sum of squares of the distance from a given point  $(x_i, u_i)$ . The function  $p(x)$  is called the fitting function or the least square solution. The method of finding the unit of the fitting function  $p(x)$  becomes the minimum of curve fitting. Two multiplication.

Analyze the jitter caused by the movement of the electronic imaging device from the kinematics of the electronic imaging device [10]. The cause of the jitter is the drastic change of speed, that is, the change of acceleration. The visual effect caused by pure uniform acceleration movement is smooth, and the curve is synthesized through the basis function. The method, using control points to determine and modify the characteristics of the curve, the expression is:

$$p(t) = \sum_{i=0}^n p_i N_{i,k}(t) \quad (16)$$

Among them,  $p_i (i = 0, 1, \dots, n)$  is the vertex of the control polygon, and  $N_{i,k}(t) (i = 1, 2, \dots, n)$  is the K-order basis function.

Assuming that the current electronic imaging equipment is at time  $k$ , the state prediction equation of the electronic imaging equipment established to predict the current state during the operation of the equipment is:

$$S(k|k-1) = \Phi \cdot S(k-1|k-1) \quad (17)$$

In the formula,  $S(k|k-1)$  is the predicted result, and  $S(k-1|k-1)$  is the optimal result at the last running time.

In order to update the electronic imaging equipment, the covariance matrix  $p$  of  $S(k|k-1)$  needs to be predicted, and the established equation is as follows:

$$P(k|k-1) = \Phi \cdot P(k-1|k-1)\Phi^T + Q \quad (18)$$

where:  $p(k|k-1)$  is the covariance matrix corresponding to  $S(k|k-1)$ ,  $p(k-1|k-1)$  is the covariance matrix corresponding to  $S(k-1|k-1)$ , and  $\Phi^T$  represents the transposed matrix of  $\Phi$ , that is, the prediction of the state of the electronic imaging device.

Now the predicted result of the filling is obtained, and then combined with the predicted value and the measured value, the optimal estimated value  $S(k|k)$  of the current state of the electronic imaging device is obtained. The updated equation is:

$$S(k|k) = \Phi \cdot S(k|k-1) + K_g(k) \cdot (Z(k) - H \cdot \Phi \cdot S(k|k-1)) \quad (19)$$

where:  $K$  represents the Kalman gain.

The size of the jitter component is the difference between the corresponding components of the global motion and the intentional motion. For the current coordinate  $k$  to be processed, the original cumulative motion vector is  $Z_{raw}(k)$ , and the smoothed motion vector after filtering is  $Z_{Kal}(k)$ , then the jitter parameter for generation compensation is  $Z(k) = Z_{Kal}(k) - Z_{raw}(k)$ , The rotation angle  $\theta(k)$ , horizontal and vertical compensation parameters  $dx(k)$  and  $dy(k)$  are obtained. In addition, the scale factor  $x(k)$  to be compensated is the product of the reciprocal scale factors between all points from the starting point. The calculation method is as follows:

$$s(k) = \prod_{i=1}^k \frac{1}{s(i)} \quad (20)$$

A typical electronic imaging device jitter and blur model in the spatial domain can be expressed as:

$$g(x, y) = f(x, y) \otimes k(x, y) + n(x, y) \tag{21}$$

Among them:  $\otimes$  is the convolution operator,  $f(x, y)$  is the potential original coordinates,  $k(x, y)$  is the fuzzy kernel of the point spread function, and  $n(x, y)$  is the noise.

Perform direct inverse filtering of the electronic imaging device to restore it. In this step, the estimated value  $f$  of the Fourier transform of the original electronic imaging device  $\hat{F}(u, v)$  is calculated by dividing the Fourier transform of the degraded device  $g$  by the Fourier transform of the degradation function  $h$ , As shown in the following formula:

$$\hat{F}(u, v) = G(u, v)/H(u, v) \tag{22}$$

However, if noise is added to the degradation function, the result of the direct inverse filtering will be very poor. Therefore, the degradation function of the equipment degradation restoration model is replaced with  $K(u, v)$ , and the estimated value optimization formula is:

$$\hat{F}(u, v) = F(u, v) + N(u, v)/H(u, v) \tag{23}$$

The above formula shows that even if the degradation function is known, the leaf cannot accurately remove the jitter that exists in the electronic imaging device [11]. In this regard, the electronic imaging equipment is subjected to non-linear iterative processing, so that the number of iterations of the de-shake operation of the equipment reaches the maximum, as shown in the following formula:

$$\hat{f}_{k+1}(x, y) = \hat{f}_k(x, y) [k(-x, -y) \otimes g(x, y)/k(x, y) \otimes \hat{f}_k(x, y)] \tag{24}$$

where:  $\hat{f}$  is the estimated value of the degenerate function. After each iteration, the band that terminates when the  $MSE$  is a constant is obtained, and the likelihood probability is obtained as:

$$p(g|f) = \prod_x \frac{(f \otimes k)(x)^{g(x)} \exp\{-(f \otimes k)(x)\}}{g(x)} \tag{25}$$

where:  $k$  is PSF and  $(f \otimes k)(x)$  is Poisson distribution. In order to obtain accurate target device characteristics, the likelihood probability  $p(g|f)$  must be maximized, and the following energy equation must be minimized to obtain the maximum likelihood solution:

$$f^* = \arg \min_f E(f) \tag{26}$$

When the derivative of  $E(f)$  and  $k$  is taken, the result of the de-jitter of the electronic imaging device is obtained, and the expression is:

$$f^{l+1} = f^l \left[ k^* \otimes \frac{g}{f^* \otimes k} \right] \quad (27)$$

where:  $k^*$  is the adjoint matrix of  $k$ , and  $l$  is the number of iterations.

## 4 Analysis of Experimental Results

CMV300 of CMOS company is selected as the photoelectric platform sensor, and the effectiveness of this method is verified by comparing with the high-precision time jitter measurement method based on self reference source.

### 4.1 Experimental Indicators

- (a) Peak signal-to-noise ratio method PSNR. What it reflects is the peak signal-to-noise ratio of the reference image frame and the current image frame. PSNR can be used as a quality factor for assessing image quality and is defined as follows:

$$PSNR(s_1, s_0) = 10 \log \frac{255^2}{MSE(s_1, s_0)} \quad (28)$$

In the formula, PSNR is an index for evaluating the standard degree of stabilization algorithm. The larger the PSNR value, the more stable the electronic imaging device, and the better the de-shake effect. If the motion vector of the two frames is zero, the PSNR value should be infinite.

- (b) The mean square error MSE is; the sum of squared differences of each pixel between two frames of images. It reflects the speed of the image sequence change and the magnitude of the change. The smaller the MSE, the more stable the electronic imaging device and the better the de-shake effect. MSE is defined as follows:

$$MSE(S_1, S_0) = \frac{1}{M \cdot N} \sum_{i=1}^N \sum_{j=1}^M (S_1(i, j) - S_0(i, j))^2 \quad (29)$$

In the formula,  $S_1(i, j)$  and  $S_0(i, j)$  respectively represent the gray value of the pixel at the point  $(i, j)$  of the current frame and the reference frame after the image sequence is compensated, and  $M$  and  $N$  represent the pixel area of the current frame and the reference pixel frame. The smaller the value of  $MSE(S_1, S_0)$  is, the higher the degree of overlap between the two images is. If the two images are completely overlapped, then  $MSE(S_1, S_0)$  should be zero.

### 5 Analysis of Experimental Results

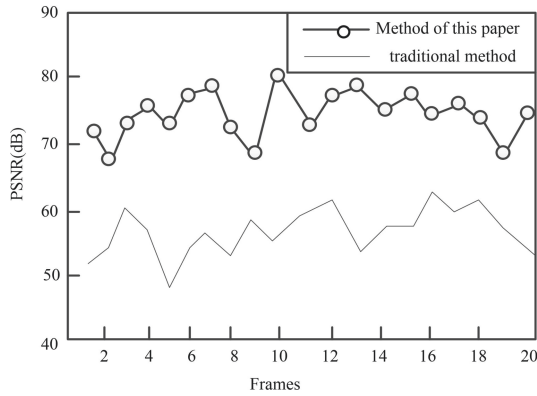


Fig. 1. PSNR comparative analysis

It can be seen from Fig. 1 that when the traditional method is used, the PSNR value is about 56.84 dB and the fluctuation is large. When the method in this paper is used for de-jitter processing, the PSNR value is about 72.64 dB, the fluctuation is small, and it is 15.8 higher than the traditional method. dB, indicating that the more stable the electronic imaging equipment adopts the method in this paper, the smaller the deviation, the better the de-jitter effect.

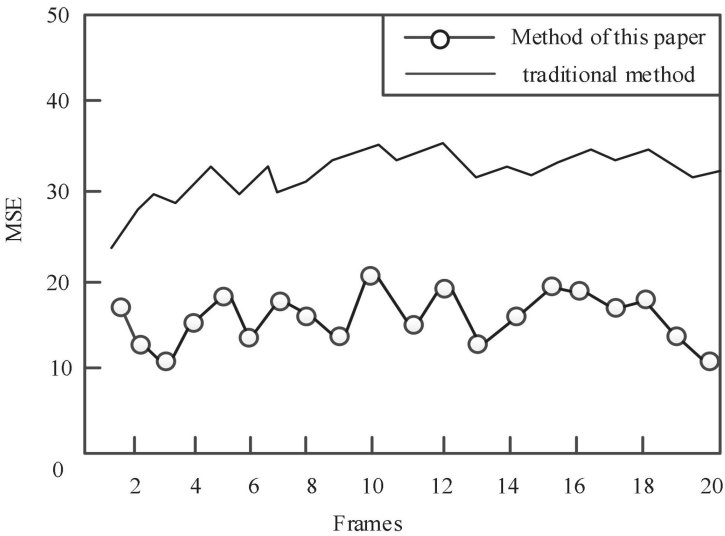


Fig. 2. Comparative analysis of mean square error MSE

It can be seen from Fig. 2 that when the traditional method is used, the MSE value is about 30.64 and the fluctuation is large. When the method in this paper is used for de-jitter processing, the MSE value is about 12.54, the fluctuation is small, and it is 18.1 lower than the traditional method. When the electronic imaging equipment adopts the method in this paper, the more stable and the smaller the deviation, the better the de-shake effect.

The experimental results show that the method in this paper compares with the traditional methods, and verifies the effectiveness of the de dithering method for electronic imaging equipment. The fluctuation of PSNR value and MSE value of this method is small, and the deviation is small. It shows that the de dithering effect of this method is better, and it can be effectively applied to the image blur problem caused by electronic imaging equipment in dynamic environment, It is of great significance to electronic imaging technology.

## 6 Conclusion

The article proposes an electronic imaging device de-shake method based on feature point matching. By acquiring device motion characteristics, using matching feature windows to achieve feature point matching of the motion path of the electronic imaging device, and determining and modifying the motion path according to the base function synthesis curve Curve characteristics, thereby optimizing the path estimation value, thereby improving the de-shake effect of the electronic imaging device. This method solves the problem of image blurring caused by the electronic imaging device in a dynamic environment to a certain extent. The results are obtained through experimental analysis. When the method in this paper is used for the de-jitter processing of electronic imaging equipment, the PSNR value is about 72.64 dB and the MSE value is about 12.54, which are better than traditional methods. However, due to the limitation of research conditions, the time-consuming situation of de-shake electronic imaging equipment needs to be further considered in future research, so as to enhance the efficiency of the de-shake method of electronic imaging equipment and improve the practical application effect of electronic imaging equipment.

## References

1. Fan, H., Qianjin, Z., Liu, L., et al.: Analysis and simulation of noise characteristics of focal plane array infrared imaging equipment. *Laser Infrared* **49**(07), 855–860 (2019)
2. Xu, P., Liu, Y., Zhang, J.: High-precision time jitter measurement based on self-reference source. *Opt. Precis. Eng.* **28**(11), 42–49 (2020)
3. Zheng, E., Cheng, Y., Lin, J.: Dense 3D reconstruction using debounce fuzzy algorithm. *Comput. Eng. Appl.* **54**(01), 217–223 (2018)
4. Xu, H., Li, S., Ji, Y., et al.: Research on panoramic camera image stitching method based on feature point matching. *Trans. Chin. Soc. Agri. Mach.* **50**(S1), 157–165 (2019)
5. Ma, Z., Mu, P., Dai, S.: Video stabilization algorithm based on feature point matching. *Packag. Eng.* **387**(21), 238–243 (2018)



6. Yang, Z., Li, Z., Zhang, G., Wen, D., Zhu, X.: Research on calculation method of actual photon number in ultraviolet imaging based on image height of electrical equipment. *High Volt. Technol.* **45**(zk), 111–116 (2019)
7. Liu, H., Li, G., Jin, G., et al.: Jitter error compensation algorithm for MIMO down-view array SAR array. *J. Survey. Mapp. Sci. Technol.* **35**(02), 175–181 (2018)
8. Li, R., Xiao, X.: Simulation of screening feature points by dithering method. *Sens. Microsyst.* **38**(05), 36–39 (2019)
9. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective Template Update Mechanism in Visual Tracking with Background Clutter. *Neurocomput.* online first (2020). <https://doi.org/10.1016/j.neucom.2019.12.143>
10. Liu, S., Liu, X., Wang, S., Muhammad, K.: Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT assisted complex environment. *Neural Comput. Appl.* **33**(4), 1055–1065 (2021)
11. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **7**(4), 1895–1917 (2020). <https://doi.org/10.1007/s40747-020-00161-4>



# An Incomplete License Plate Image Intelligent Recognition System Based on the Generated Counter Network

Mi Meng<sup>1(✉)</sup>, Chun-hu He<sup>2</sup>, and Xiao-jing Qi<sup>1</sup>

<sup>1</sup> Chongqing Telecommunication Vocational College,  
Chongqing 402274, China  
mengmi4545@163.com

<sup>2</sup> Chongqing Toramon Technology Co Ltd., Chongqing 401121, China

**Abstract.** In order to solve the problems of high recognition accuracy and long time-consuming recognition in the traditional incomplete license plate image recognition system, the paper proposes an incomplete license plate image intelligent recognition system based on a generative confrontation network. According to the composition of the recognition system, the framework of the intelligent recognition system for incomplete license plate images is designed. The hardware platform of the system is the S3C6410 embedded development board, which is based on Samsung's ARM11 processor chip, and the CPU is based on the ARM111176JZF-S core design. A powerful multimedia processing unit is integrated inside, which supports hardware encoding and decoding of video files in formats such as Mpeg4, H.264/H.263, and can be output to LCD and TV display at the same time. The 6410 uses a high-density 6-layer board design, which integrates 256M DDR RAM, 256M/1 GB SLC Nand Flash memory. The system software uses a generative confrontation network to repair the incomplete license plate image, and combines the template matching method to complete the license plate image recognition. Experimental results show that the designed intelligent recognition system has high recognition accuracy and can effectively reduce the time-consuming recognition.

**Keyword:** Generative counter measure network · Incomplete license plate image · Intelligent recognition

## 1 Introduction

License plate image intelligent recognition system is a special computer vision system, which takes the car photo as a specific target. It is one of the important research topics of computer vision and pattern recognition in the field of intelligent transportation. It can be widely used in traffic flow detection, traffic control and guidance, intelligent community management, automatic toll collection without stopping, red light running and other illegal vehicle monitoring, vehicle security and other fields, and has broad application prospects [1–3].

With the rapid development of China's economy, the scale of highway network is expanding, and effective traffic management technology is extremely necessary and

urgent. The traffic management mode mainly includes artificial traffic management and intelligent transportation system management. Because the latter can save a lot of manpower and material resources, intelligent transportation system has become the main direction of traffic management development. Intelligent transportation system integrates the traffic management system effectively by means of control, communication, computer and optics, and realizes the efficient, accurate and comprehensive transportation management. License plate recognition system, as an important tool of intelligent transportation system, plays an important role in intelligent traffic management. With the rapid development of digital image processing technology, license plate recognition system has been continuously optimized. The main task of license plate recognition system is to process and analyze the license plate image collected by camera, so as to complete the recognition of vehicle license plate information [5–7].

License plate is the unique sign of vehicles, and its recognition is conducive to the accurate and efficient management of transportation. Since the 21st century, the domestic license plate recognition technology has developed rapidly. Researchers have proposed a variety of license plate location and recognition methods based on the characteristics of Chinese license plate, but most license plate recognition systems have many limitations, such as: the resolution of license plate image must be high and clear as a whole; There should be no text advertisements around the license plate; The license plate shall not be seriously tilted or incomplete. Therefore, it is of great significance to study a management system which can locate and recognize license plate efficiently and accurately.

Therefore, this paper proposes an intelligent recognition system of incomplete license plate image based on generative countermeasure network. The overall design idea of the system is: Based on S3C6410 embedded development board, the hardware platform of the system is designed to realize the real-time transmission and display of license plate image, and the software part of the system uses generative countermeasure network to repair the incomplete license plate image, combined with the template matching method to complete the license plate image recognition. Finally, the effectiveness of the designed system is verified by experiments.

## **2 Design of Intelligent Recognition System for Incomplete License Plate Image**

The completed license plate recognition system is generally composed of vehicle perceptron, image acquisition, license plate character recognition, recognition result output and other parts [8].

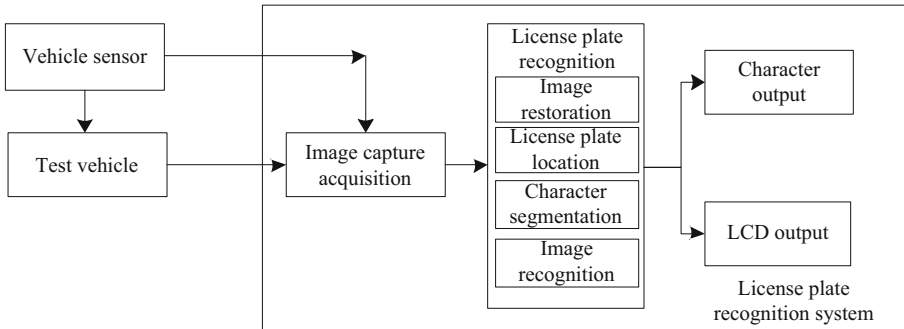
Vehicle sensor is to sense whether the vehicle has reached, whether to capture and obtain the image. There are external trigger devices commonly used, such as: ground sensor coil sensor, which triggers the signal when the vehicle comes and sends it to the system to capture the license plate image. In addition, there are other trigger methods, such as video stream trigger.

Image acquisition, this process is mainly when there is a trigger signal, the use of cameras and other image acquisition devices to capture the vehicle image, stored in the license plate recognition system, easy to further identify [9–11].

License plate recognition, the main function is to process the captured license plate image, including image binary, license plate positioning, character segmentation and other parts. The special feature of the system is to use the generated counter network to recover incomplete images, and finally output the license plate recognition results through character recognition [12, 13].

The output of recognition results, the identified string shall be output to other systems or devices through other interfaces for other systems.

The system framework structure is shown in Fig. 1.



**Fig. 1.** Frame structure of license plate recognition system

The working principle of the system is: when the vehicle passes through the monitoring area, the vehicle monitoring unit is triggered, the signal is transmitted to the camera, the camera collects the vehicle license plate image and transmits the image to the recognition processing unit for license plate recognition, and the recognition results are displayed on the LCD.

## 2.1 System Hardware

The hardware platform of the system is the S3C6410 embedded development board, which is based on Samsung's ARM11 processor chip, the CPU is based on the ARM111176JZF-S core design, and a powerful multimedia processing unit is integrated inside. Supports hardware encoding and decoding of video files in formats such as Mpeg4, H.264/H.263, and can be output to LCD and TV display at the same time. The 6410 uses a high-density 6-layer board design, which integrates 256M DDR RAM, 256M/1GB SLC Nand Flash memory. It adopts 5V power supply, realizes various core voltage conversions necessary for the CPU on the motherboard, and also has a professional reset chip. It has the characteristics of high performance, rich interfaces, small size and low power consumption. It is suitable for the secondary development of various mobile handheld devices. The camera uses a USB camera to obtain license plate images, and the LCD display is a 4.3-in. touch screen.

The software operating environment for license plate image recognition is an embedded Linux operating system, and the development environment is Qt. Qt is used

for visual programming. At the same time, the machine vision library OpenCV is used for image processing when processing images.

## 2.2 System Software

### Incomplete Image Restoration Based on the Generated Counter Network

The generative confrontation network was designed and proposed in 2014. It is a generative model. Its core idea is a two-person zero-sum game. The generative confrontation network model is composed of a generative network  $G$  and a discriminant network  $D$ . The generative network  $G$  generates new data by continuously learning the probability distribution of real data and using the learned distribution model. The role of the discriminant network  $D$  is to optimize the mutual parameters of the real data and the generated network without much prior knowledge of the image. Improve their respective generation judgment ability, and finally make the data generated through the generation network very close to the real data. The original model framework of the generative confrontation network is shown in Fig. 2.

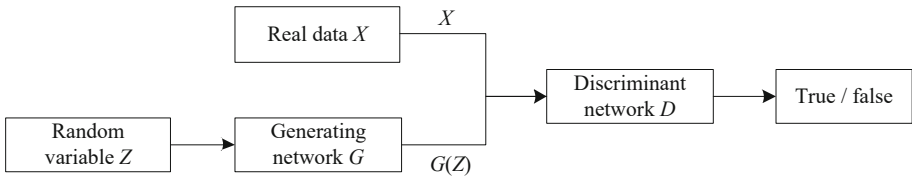


Fig. 2. GAN structure

The random noise vector  $Z$  is input to generate the network  $G$  and output  $G(z)$ . When the input data of the discrimination network  $D$  is  $G(z)$ ,  $D$  outputs 0. The purpose of network countermeasure training is to make  $G(z)$  close to the real data distribution  $P_{date}$ . After continuous confrontation training and iterative optimization, when  $D$  is unable to identify the data source, it is considered that the generated network has learned the real data distribution [14–16].

The loss function of generating network  $G$  is as follows:

$$\min_G(D, G) = E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{1}$$

The loss function of network  $D$  is as follows:

$$\max_D V(D, G) = E_{x \sim p_{date}(x)}[\log(1 - D(x))] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{2}$$

The total loss function of the model is as follows:

$$\min_G \max_D (D, G) = E_{x:p_{date}(x)}[\log D(x)]J + E_{z:p_z(z)}[1 - D(G(z))] \tag{3}$$

In the formula,  $E$  is the function of distribution function,  $p_{date}$  is the real data distribution,  $p_z$  is the noise distribution.

Incomplete license plate image restoration is an attempt to use the prior knowledge of the degradation process to restore the original appearance of the degraded image, that is, according to the causes of degradation, analyze the environmental factors causing degradation, establish the corresponding data model, and restore the image along the reverse process of the degradation of license plate image.

In practical application, it is usually assumed that the transmission system is a linear system, and the original image  $f(x, y)$  passes through the system  $h(x, y)$ .  $h(x, y)$  is a system function obtained by integrating all degradation factors, which is called impulse response or point spread function in urban-rural system. The structure of the basic degradation model is shown in Fig. 3.

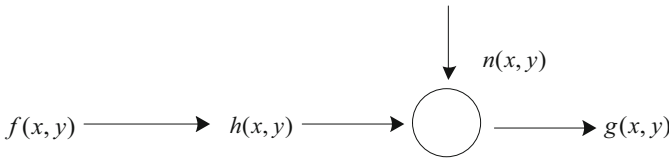


Fig. 3. Degradation model

In Fig. 3,  $g(x, y)$  is the actual degraded image, and  $n(x, y)$  is the incomplete noise model.

According to the degradation model shown in Fig. 3, the problem of image restoration can be regarded as a process of estimating the original image  $f(x, y)$  and making the estimated image as close as possible to the real image  $f(x, y)$  according to the obtained incomplete image  $g(x, y)$ , the prior knowledge of  $h(x, y)$  and some statistical properties of incomplete noise  $n(x, y)$ . If there is no prior knowledge of PSF, incomplete image restoration is more difficult.

According to the image degradation block diagram, the degradation model can be expressed as:

$$g(x, y) = f(x, y) * h(x, y) + n(x, y) \tag{4}$$

In practical applications, all digital images are processed, and the above formula is expressed in discretized form as follows:

$$g(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n)h(x - m, y - n) + n(x, y) \tag{5}$$

In the formula,  $x = 0, 1, 2, \dots, M - 1, y = 0, 1, 2, \dots, N - 1$ .

In practical applications, image integrity restoration is often carried out without knowing the PSF, and blind deconvolution restoration is proposed in this context. Blind deconvolution restoration is an image inpainting method that uses the original incomplete image to estimate PSF and complete image at the same time.

It can be seen from the above formula that deconvolution must be performed to solve  $f$ . Using the method of least squares constraint, constrain  $f$  and  $h$  at the same time, and choose  $r(f) + r(h)$  as the penalty function, which can be obtained by the above formula:

$$n = g - h * f \tag{6}$$

and:

$$E\left(\int n^2 dx\right) = \sigma^2 \tag{7}$$

and there are:

$$H_1(u) = \int |\nabla u|^2 dx dy \tag{8}$$

So:

$$|h * f - g|^2 = E\left[\int (h * f - g)^2 dx\right] \approx E\left(\int n^2 dx\right) = \sigma^2 \tag{9}$$

Then the image integrity repair problem can be attributed to the following minimum constraint problem:

$$\min[\partial_1 r(f) + \partial_2 r(h)] \tag{10}$$

When

$$\|h * f - g\|^2 = \|n\|^2 \approx \sigma^2 \tag{11}$$

The corresponding Lagrangian form is:

$$\min L(f, h) = \min\left[\|h * f - g\|^2 + \alpha_1 r(f) + \alpha_2 r(h)\right] \tag{12}$$

The  $\alpha_1$  and  $\alpha_2$  are coefficients after the Lagrangian multiplier. The key to the problem is the selection of  $r(f)$  and  $r(h)$ . If  $\nabla$  is gradient operator in  $H_1$  rule, the above equation can be written as follows:

$$\begin{aligned} \min L(F, H) &= \min \left[ \|h * f - g\|^2 + \alpha_1 \int |\nabla f|^2 dx dy + \alpha_2 \int |\nabla h|^2 dx dy \right] \\ &= \min \left[ (h * f - g)^2 dx dy + \alpha_1 \int |\nabla f|^2 dx dy + \alpha_2 \int |\nabla h|^2 dx dy \right] \\ &= \int \min \left[ (h * f - g)^2 + \alpha_1 |\nabla f|^2 + \alpha_2 |\nabla h|^2 \right] dx dy \end{aligned} \tag{13}$$

The partial derivative of the part within the integral sign in the above formula can be obtained as follows:

$$0 = h(-x, -y) * (h * f - g) - 2\alpha_1 \Delta f \tag{14}$$

$$0 = f(-x, -y) * (h * f - g) - 2\alpha_2 \Delta h \tag{15}$$

In the formula,  $\Delta$  is Laplace operator. In the practical application, the processing is all digital images, and the corresponding discretization form is as follows:

$$H * (Hf - g) - \alpha_1 \Delta f = 0 \tag{16}$$

$$F * (Fh - g) - \alpha_2 \Delta h = 0 \tag{17}$$

After finishing the above formula, we can get:

$$[H * H + \alpha_1(-\Delta)]f = H * g \tag{18}$$

$$[F * F + \alpha_2(-\Delta)]h = F * g \tag{19}$$

Let  $F(\xi_x, \xi_y)$  be the Fourier transform of  $f(m, n)$ ,  $h(\xi_x, \xi_y)$  be the Fourier transform of  $h(m, n)$ ,  $G(\xi_x, \xi_y)$  be the Fourier transform of  $g(m, n)$ , and  $R(\xi_x, \xi_y)$  be the Fourier transform of  $(-\Delta)$ , then the frequency domain form of the above two equations is:

$$F(\xi_x, \xi_y) = \frac{H * (\xi_x, \xi_y)G(\xi_x, \xi_y)}{|F(\xi_x, \xi_y)|^2 + \alpha^2(\xi_x, \xi_y)} \tag{20}$$

$$H(\xi_x, \xi_y) = \frac{F * (\xi_x, \xi_y)G(\xi_x, \xi_y)}{|F(\xi_x, \xi_y)|^2 + \alpha_2 R(\xi_x, \xi_y)} \tag{21}$$



Here is an empirical formula for  $R$ :

$$R = (\xi_x, \xi_y) = 4 - 2 \cos(2\pi\xi_x/M) - 2 \cos(2\pi\xi_y/M) \tag{22}$$

In order to improve the accuracy of restoration, the iterative method is adopted. First fix  $h^k$ , solve  $f^{(k+1)}$ , and then solve  $h^{(k+1)}$  according to  $f^{(k+1)}$ . In this way, the iteration formula is as follows:

$$F^{(k+1)}(\xi_x, \xi_y) = \frac{H^{k^m}(\xi_x, \xi_y)G(\xi_x, \xi_y)}{|H^k(\xi_x, \xi_y)|^2 + \alpha_1 R(\xi_x, \xi_y)} \tag{23}$$

$$H^{(k+1)}(\xi_x, \xi_y) = \frac{F^{(k+l)^m}(\xi_x, \xi_y)G(\xi_x, \xi_y)}{|H^{(k+l)}(\xi_x, \xi_y)|^2 + \alpha_2 R(\xi_x, \xi_y)} \tag{24}$$

Through the above formula, the incomplete image restoration is completed by generating the anti network.

**License Plate Character Recognition**

The purpose of character recognition is to transform image into text information, which is an important branch of pattern recognition. As the last part of license plate recognition system, character recognition accuracy directly affects the performance of the whole license plate recognition system. In addition, the accuracy of character recognition method is seriously affected by the result of character segmentation. If the method of character recognition is based on machine learning, its recognition accuracy is also affected by the completeness of the training set.

The main steps of the character recognition algorithm based on template matching are:

First, compare the characteristics of the character to be recognized with the characteristics of the standard character; then, find the most similar expression form, then the recognition result is the character corresponding to this expression form.

The essence of the character recognition method based on template matching is to calculate the similarity between the training sample and the test sample. The category of the training sample with the greatest similarity to the test sample is the category of the test sample. Using character characteristics and matching principles to recognize character images is to judge test samples and training in a classification model. A character recognition algorithm for template matching will be described in detail below.

Suppose that the function  $f(x, y)$  represents the input character, the function  $F(x, y)$  represents the standard template, and the output result is  $T(x, y)$ . Now, using the correlation quantity  $x_1$  and  $x_2$  to represent the random variable, the output result is as follows:

$$T(x_1 - x_2, y_1 - y_2) = \iint f(x, y)F(x + (x_1 - x_2), y + (y_1 - y_2))dxdy \quad (25)$$

When  $x_1 = x_2, y_1 = y_2,$

$$T(0, 0) = \iint f(x, y)F(x, y)dxdy \quad (26)$$

When  $f(x, y) = F(x, y),$

$$T(0, 0) = \iint f(x, y)f(x, y)dxdy \quad (27)$$

That is, the correlation function of the test sample, and  $T(0, 0) \geq T(x, y)$ .  $T(0, 0)$  is the main peak of  $T(x, y)$ , and there will be some sub peaks which are not equal to the main peak in other characters. Then, the character category can be determined by setting the threshold.

The size of the test image and the training image is normalized, and then the similarity between the template and the character image is calculated. Suppose that the image to be detected is represented as  $F(m, n)$ , the image sub image to be detected is represented as  $F_{xy}(m, n)$ , the coordinates of the upper left corner pixel of the sub image in the image are  $(x, y)$ , and the template image is  $T(m, n)$ , then the correlation operator  $R(x, y)$  is:

$$R(x, y) = \frac{\sum_{m=1}^M \sum_{n=1}^N F_{xy}(m, n) * T(m, n)}{\sqrt{\sum_{m=1}^M \sum_{n=1}^N (F_{xy}(m, n))^2 \sum_{m=1}^M \sum_{n=1}^N (T(m, n))^2}} \quad (28)$$

The best matching template is the corresponding template with the largest correlation operator. Since the original image is a binary image, the above formula can be simplified to:

$$D(x, y) = \sum_{m=1}^M \sum_{n=1}^N F_{xy}(m, n) \oplus T(m, n) \quad (29)$$

In the formula, the smaller the value of  $D(x, y)$ , the better the repair effect of incomplete license plates.

### 3 Experimental Verification

In order to verify the practical application performance of the proposed intelligent recognition system of incomplete license plate image based on generative counter-measure network, simulation and comparative verification experiments are carried out.

#### 3.1 Experimental Scheme

System experiment environment:

Hardware: Intel Core i5@2.6 GHz, RAM 8 GB.

Software: Windows 10 64-bit operating system, Microsoft Visual C++ 2017 (v141), Qt 5.10.1 (Open Source License).

Experimental data: There are few publicly available incomplete license plate image databases. In order to better carry out simulation experiments, 3000 pairs of images have been collected from the network image library and the GoPro data set to form a data set. Each pair of images contains an incomplete image and a corresponding complete image. Therefore, the number of images used in the experiment is 1500, and the resolution size of the image pair is set to  $256 \times 256$ . The sample image is shown in Fig. 4.



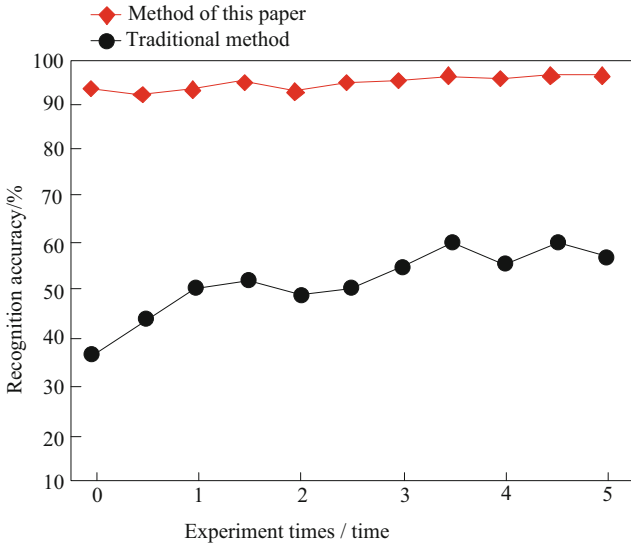
Fig. 4. Sample image

Experimental scheme: Taking the incomplete license plate recognition accuracy and recognition time as the experimental comparison index, this method is compared with the traditional method.

#### 3.2 Analysis of Experimental Results

##### Recognition Accuracy

Because the license plate images are incomplete, it is necessary to verify the recognition accuracy of incomplete license plate images. The comparison results of the recognition accuracy of incomplete license plate image between the proposed method and the traditional method are shown in Fig. 5.



**Fig. 5.** Comparison of image recognition accuracy of incomplete license plates

The results of the comparison of the image recognition accuracy of incomplete license plate shown in Fig. 5 show that the recognition accuracy of this method is basically kept at more than 90%. Although the traditional method can keep the fluctuation rising state, the maximum recognition accuracy is still not 70%.

**Identification Time**

In the process of recognizing incomplete license plate image, it is necessary to restore the incomplete license plate image, so it is necessary to verify the recognition time of incomplete license plate image. The results of time-consuming comparison between the proposed method and the traditional method for recognition of 1500 license plate images are shown in Table 1.

**Table 1.** Comparison results of incomplete license plate image recognition

Experiment times/time	Identification time/s	
	Method of this paper	Traditional method
1	1423	1898
2	1412	1941
3	1429	1967
4	1408	1983
5	1431	2011
Mean value	1420.6	1960

From the time comparison results of incomplete license plate image recognition shown in Table 1, it can be seen that when there are 1500 incomplete images to be recognized, the average recognition time of this method is 1420.6s, while the average recognition time of traditional method is 1960s. Therefore, this method can effectively reduce the recognition time.

## 4 Conclusion

License plate recognition is one of the research hotspots in the field of modern computer vision and pattern recognition. With the wide application of intelligent transportation system and the rapid growth of vehicle data, it is more and more important to use license plate recognition system for vehicle management and road dredging. License plate recognition is paid more and more attention by domestic and foreign scientific research institutions. In the context of the popularity of mobile office and mobile commerce, the mobile license plate recognition is also on the agenda. However, the current license plate recognition system used in complex background, its recognition accuracy is not high enough. Therefore, it is of great significance for traffic management to develop a license plate recognition system with perfect functions and high recognition accuracy. In this paper, the incomplete license plate image recognition system based on generative countermeasure network is comprehensively designed. The recognition performance of the designed system is verified by experiments. The method has high recognition accuracy and efficiency in incomplete license plate image recognition.

**Fund Projects.** 1. Project name: incomplete license plate recognition study based on generative adversarial network, project number: KJQN201905503  
 2. Project name: The work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission, project number: KJZD-K201801901  
 3. roject name: research and design of intelligent manufacturing cloud platform system based on Internet of things, project number: KJQN201801902

## References

1. Li, H., Hao, P., Liao, Y., et al.: Design of license plate detection and recognition system. *Comput. Tech. Dev.* **30**(07), 150–153+159 (2020)
2. Zhang, X., Jin, J.: End-to-end license plate detection and recognition system based on quantized neural network. *Transducer Microsys. Technol.* **39**(12), 108–110 (2020)
3. Zhao, W., Zhang, N.: Simulation of license plate image enhancement algorithm in severe hazy weather. *Comput. Simul.* **36**(03), 213–217 (2019)
4. Zheng, G., Wu, H.: Method of license plate location based on MSER feature and edge projection. *Comput. Eng. Des.* **40**(01), 241–244 (2019)
5. Liu, S., Liu, G., Zhou, H.: A Robust parallel object tracking method for illumination variations. *Mob. Netw. Appl.* **24**(1), 5–17 (2018). <https://doi.org/10.1007/s11036-018-1134-8>
6. Duan, B., Fu, X., Jiang, Y., et al.: Lightweight blurred car plate recognition method combined with generated images. *J. Image Graph.* **25**(9), 1813–1824 (2020)

7. Li, M., Zhang, J., Wu, S., et al.: License plate image deblurring algorithm based on RANSAC transform. *Transducer Microsyst. Technol.* **39**(02), 158–161+165 (2020)
8. Hu, X., Xu, X., Bai, Y., et al.: License plate image recognition algorithm based on shape context. *Electronic Des. Eng.* **27**(09), 131–135 (2019)
9. Yao, X., Wang, X., Wang, S.H., Zhang, Y.D.: A comprehensive survey on convolutional neural network in medical image analysis. *Multimedia Tools Appl.* **24**, 1–45 (2020). <https://doi.org/10.1007/s11042-020-09634-7>
10. Rongmei, Z., Qi, Z., Bin, C.: An improved license plate recognition algorithm based on LeNet-5 convolutional neural network. *Sci. Technol. Eng.* **20**(12), 4775–4779 (2020)
11. Zhang, F., Wang, X., Hao, X.: Intelligent Vehicle character recognition based on edge feature. *Autom. Instrum.* **28**(06), 17–20+26 (2020)
12. Feige, Z., Kai, L., Shaokang, Z., et al.: Application of MATLAB in digital image processing. *Comput. Technol. Dev.* **29**(11), 222–226 (2019)
13. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. *Multimedia Tools Appl.* **76**(4), 5787–5802 (2014). <https://doi.org/10.1007/s11042-014-2408-1>
14. Hu, J., Li, X.: Cross-modal pedestrian re-identification based on generative confrontation network. In: 2021 International Conference on Electronics, Circuits and Information Engineering (ECIE) (2021)
15. Ma, W., Zhang, Y., Guo, J., et al.: Abnormal traffic detection based on generative adversarial network and feature optimization selection. *Int. J. Comput. Intell. Syst.* **14**(1), 26–32 (2021)
16. Weng, Y., Zhou, H., Wan, J.: Image inpainting technique based on smart terminal: a case study in CPS ancient image data. *IEEE Access* **56**(19), 131–136 (2019)

## Author Index

- An, Jing I-53
- Bai, Yang I-53
- Bi, Shuhui II-440
- Bian, Hai-hong II-461
- Cai, Chen II-386
- Cao, Jing II-352, II-394
- Cen, Cai-cun I-212
- Chai, Hui I-464, II-115, II-129
- Chen, Bao-ren I-3, I-40
- Chen, Lu II-508
- Chen, Shu-Wen II-497, II-508
- Chen, Yong-ming I-212, II-140
- Chu, Hao-nan II-71, II-88
- Cui, Dai I-344
- Cui, Jinrong II-313, II-321
- Duan, Xiu-juan II-153, II-171
- Fan, Hua I-171, II-268
- Feng, Jidong II-358, II-394
- Gai, Yu-hua I-200
- Gao, Bingbing II-419
- Gao, Mingyang I-272
- Gao, Peng I-15
- Gao, Yanli II-402
- Geng, Renkang II-333, II-352, II-368
- Gong, Li-kuan I-27, II-461
- Gu, Xiao-Wei II-497, II-508
- Guo, Hui I-212
- Guo, Xin-Ze I-447
- Han, Bing-bing I-187, II-58
- Han, Hai-yun I-187, II-58
- Han, Lin II-358
- He, Chun-hu I-361, I-570
- He, Dan-kang I-433, I-476
- Hong, Dan-ke I-3, I-40
- Hu, Guoxing II-386
- Hu, Zi-Peng I-447
- Hu, Zi-peng I-136
- Huang, Cheng II-321
- Huang, Guan-jin I-15
- Huang, Hong-yu I-526, I-542
- Huang, Qiu-jiao I-433, I-476
- Ji, Shaokang II-402
- Jiang, Di I-344
- Jiang, Feng I-326
- Jin, Xue I-119
- Lei, Tao I-154, II-30
- Lei, Tengfei II-433
- Li, Chao I-136, I-447
- Li, Jia-hua I-290, I-308
- Li, Mingran II-333, II-394, II-402
- Li, Tie I-326, I-344
- Li, Wenmin II-419
- Li, Xia II-386
- Li, Xue-mei II-46
- Li, Ye-fei I-71
- Li, Zhihao II-440
- Liang, Huan-wei II-183, II-199
- Lin, Shuai II-358, II-368, II-376
- Liu, Bin-bin I-105, I-119
- Liu, Hailong II-313
- Liu, Hui I-71
- Liu, Jin-hua I-226, I-244
- Liu, Kai I-136, I-447
- Liu, Xiao-xiao I-136
- Liu, Zhi-wei I-27
- Lv, Qiu-ying I-200, II-249
- Ma, Chong-wei I-71
- Ma, Dong-bao I-464, II-46
- Ma, Liyao II-368, II-376
- Ma, Wanfeng II-341, II-394
- Ma, Xiao-gang II-3, II-19
- Meng, Mi I-361, I-570
- Meng, Xian-ying I-171
- Miao, Guang-yao I-417
- Ni, Longqiang II-419
- Niu, Qing-li I-373, II-103
- Ou, Yong-tong I-3, I-40

- Peng, Chaoda II-313  
 Pi, Jun-bo I-326
- Qi, Xiao-jing I-260, I-570  
 Qi, Xiao-jing I-559  
 Qin, Qing-huan I-526, I-542  
 Qu, Jianling II-402  
 Qu, Ming-fei II-46
- Ren, Wanjie II-386, II-394  
 Ruan, Rong-rong I-88
- Shen, Tao II-451  
 Shi, Xing-jian II-461  
 Shu, Yanxin I-272  
 Su, Yang I-200, II-249  
 Su, Zai-xing I-187  
 Sun, Bo II-71, II-88  
 Sun, Haoxuan II-341, II-358, II-376  
 Sun, Jian-mei II-153, II-171  
 Sun, Mingxu II-333, II-358  
 Sun, Xin I-373
- Tang, Jun-ci I-326, I-344  
 Tong, Xuan II-508  
 Tuo, Rui II-386
- Wan, Xiao-zheng I-464, II-46, II-115, II-129  
 Wang, Feng I-386, I-399  
 Wang, Guo-bin I-53, I-71  
 Wang, Huan-yu II-3, II-19  
 Wang, Le-le I-53, I-417  
 Wang, Li I-3, I-40  
 Wang, Miao I-326  
 Wang, Miao-geng I-15  
 Wang, Qian II-461  
 Wang, Tingting II-341  
 Wang, Wei II-419, II-479, II-488  
 Wang, Xinyao II-440  
 Wang, Yujie II-333, II-352  
 Wei, Miao I-492, I-509
- Whang, Shu-hua I-171  
 Wu, Si-Ye II-508  
 Wu, Tong-hao I-27
- Xie, Gui-xiu I-154, II-30  
 Xie, Wen-da II-216, II-233  
 Xie, Yun-xia I-492, I-509  
 Xiong, Kui I-272  
 Xu, Cai-xu I-212, II-140  
 Xu, Hong I-373, II-103  
 Xu, Junpeng II-376  
 Xu, Qinhua II-451  
 Xu, Wen-Lin I-136, I-447  
 Xu, Wen-tao I-53, I-71  
 Xu, Yuan II-394, II-451
- Yan, Fengshuo I-272  
 Yan, Feng-shuo I-417  
 Yan, Hui-Shen II-508  
 Yao, Wen-shan I-171  
 Ye, Meng I-15, I-88  
 Ying, Jie-yao I-373  
 Yu, Jian-ming I-344  
 Yue, Dong-ming I-417
- Zang, Hongyan II-433  
 Zeng, Qing-bang II-216, II-233  
 Zhang, Guo-yi I-27  
 Zhang, Hai-bo II-153, II-171  
 Zhang, Hui-ling I-417  
 Zhang, Ji-ming I-464, II-115, II-129  
 Zhang, Kun II-341, II-368, II-376  
 Zhang, Shourong II-433  
 Zhang, Song I-464, II-115, II-129  
 Zhang, Xu-hui I-15  
 Zhao, Dan II-183, II-199  
 Zhao, Huan-yu II-115, II-129  
 Zhao, Qinjun II-341, II-352, II-440, II-451  
 Zheng, Min-na I-171  
 Zhong, Fu-lian I-226, I-244  
 Zhong, Xin-hui I-3, I-40  
 Zhou, Fengguang II-451  
 Zhou, Hai-jing II-280, II-298  
 Zhou, Jian-yong I-27