Journal of Electronic Research and Application

Editor-in-Chief

Joselito A. Dolot

Lyceum of the Philippines University-Batangas, Philippines

BIO-BYWORD SCIENTIFIC PUBLISHING PTY LTD
(619 649 400)
Level 10
50 Clarence Street
SYDNEY NSW 2000
Copyright © 2024. Bio-Byword Scientific Publishing Pty Ltd.

ISSN (ONLINE): 2208-3510
ISSN (PRINT): 2208-3502

Journal of Electronic Research and Application

Focus and Scope

Journal of Electronic Research and Application is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application.

Topics covered but not limited to:

Automation
Circuit Analysis and Application
Electric and Electronic Measurement Systems
Electrical Engineering
Electronic Materials
Electronics and Communications Engineering
Power Systems and Power Electronics
Signal Processing
Telecommunications Engineering
Wireless and Mobile Communication

About Publisher

Bio-Byword Scientific Publishing is a fast-growing, peer-reviewed and open access journal publisher, which is located in Sydney, Australia. As a dependable and credible corporation, it promotes and serves a broad range of subject areas for the benefit of humanity. By informing and educating a global community of scholars, practitioners, researchers and students, it ndeavors to be the world's leading independent academic and professional publisher. To realize it, it keeps creative and innovative to meet the range of the authors' needs and publish the best of their work.

By cooperating with University of Sydney, University of New South Wales and other world-famous universities, Bio-Byword Scientific Publishing has established a huge publishing system based on hundreds of academic programs, and with a variety of journals in the subjects of medicine, construction, education and electronics.

Publisher Headquarter

BIO-BYWORD SCIENTIFIC PUBLISHING PTYLTD
Level 10
50 Clarence Street
Sydney NSW 2000
Website: www.bbwpublisher.com
Email: info@bbwpublisher.com

2024, Volume 8, Issue 1

Table of Contents

1 **Enhancing Hazardous Chemical Management in Chinese University Laboratories: Strategies for Safety and Efficiency**
Yizhe Li

7 **Research on Modulation Signal Denoising Method Based on Improved Variational Mode Decomposition**
Canyu Mo, Qianqiang Lin, Yuanduo Niu, Haoran Du

16 **Instrument Automation Control System Faults and Maintenance**
Jianwei Li

22 **Construction Technology of Lightning Protection Devices in Building Electrical Installation Engineering**
Xuhong Du

28 **The Impact of Temperature on the Electromagnet and Structural Optimization of a Solenoid Valve**
Zhixiong Tang, Zhijun Feng

36 **An Affective EEG Analysis Method Without Feature Engineering**
Jian Zhang, Chunying Fang, Yanghao Wu, Mingjie Chang

46 **Research and Application of Artificial Intelligence in Electric Guitar Performance**
Tianchang Li

51 **Analysis of Traffic Accidents Based on the Integration Model**
Yanshun Ma, Yi Shi, Yihang Song, Chenxiao Wu, Yuanzhi Liu

60 **Comprehensive Benefit Evaluation of SZ Distributed Photovoltaic Power Generation Project Based on AHP-Matter-Element Extension Model**
Shuli Jing

69 **Crossing the Achilles Heel of Algorithms: Identifying the Developmental Dilemma of Artificial Intelligence-Assisted Judicial Decision-Making**
Kexin Chen

73 Simulation-Based Investigation of Crack Formation in Forged MB15 Magnesium Alloy
Lulu Song, Dongmei Liu, Jie Yuan, Ruihuan Tian, Liying Jin

80 Attribute Reduction of Neighborhood Rough Set Based on Discernment
Biqing Wang

86 Synthesis of Ionic and Non-Ionic Modified Waterborne Polyurethane Curing Agent
Ying Qiao, Xu-dong Zhang

93 Design of Cold-Junction Compensation and Disconnection Detection Circuits of Various Thermocouples and Implementation of Multi-Channel Interfaces Using Them — A Secondary Publication
Hyeong-Woo Cha

106 A Study on the Design Method of Hybrid MOSFET-CNTFET Based SRAM – A Secondary Publication
Geunho Cho

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Enhancing Hazardous Chemical Management in Chinese University Laboratories: Strategies for Safety and Efficiency

Yizhe Li*

Northwest Normal University, Lanzhou 730070, Gansu Province, China

***Corresponding author:** Yizhe Li, 202131805514@nwnu.edu.cn

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: This paper examines the management of hazardous chemicals in Chinese university laboratories, identifying key challenges and proposing improvements. It reviews current practices and safety measures, highlighting deficiencies such as inadequate safety systems and insufficient awareness among personnel. The study emphasizes the necessity of tailored safety management systems, the integration of digital tracking technologies like Radio Frequency Identification, and enhanced safety training for staff. The proposed recommendations aim to mitigate risks and enhance laboratory safety and efficiency. In conclusion, the paper asserts that a comprehensive approach, encompassing improved management systems, technological advancements, and educational initiatives, is essential for safer chemical handling in academic research environments

Keywords: Hazardous chemicals; Laboratory safety; RFID technology; Safety management; Chinese universities

Online publication: January 18, 2024

1. Introduction

University laboratories serve as primary vessels for conducting teaching and research activities, playing a crucial role in nurturing students' innovative and practical skills. They are pivotal in the construction of "Double First-Class" universities. Laboratories in experimental disciplines often require various reagents and chemicals, including regulated chemicals. Regulated chemicals refer to those substances that, under legal and regulatory provisions, are subject to special control measures by the state to prevent their misuse in illegal activities. This category includes highly toxic substances, precursors for drug and explosive production, civilian explosives, psychotropic drugs, and anesthetics. Improper use or management of these regulated chemicals can potentially lead to significant safety incidents, adversely impacting social stability [1].

According to relevant statistics, between 2016 and 2020, China witnessed 929 chemical and hazardous chemical accidents, resulting in 1,176 fatalities. These series of safety incidents have sounded another alarm for laboratory safety management and have drawn close attention from all sectors of society to the safe

management of hazardous chemicals. It is evident from these events that accidents caused by hazardous chemicals are characterized by their unpredictability and immeasurable consequences. Laboratory accidents inevitably lead to severe repercussions, including loss of property in scientific institutions, hindrance in research progress and project timelines, and even threats to the safety of scientific personnel. Considering the diversity of laboratories, the complexity and fluidity of personnel levels, and the characteristics of hazardous chemicals, research institutions must control the process risk factors of hazardous chemical safety management before, during, and after usage. By perfecting related laboratory safety management systems, fulfilling the responsibilities of various functional departments, and enhancing the safety awareness of scientific personnel, it is possible to prevent laboratory accidents and ensure the smooth progression of research projects, thereby providing a safe research environment for scientists [2].

2. Issues in chemical management in China

2.1. Inadequate safety management systems

In recent years, to enhance laboratory safety management, the Chinese government has introduced a series of laws, regulations, and standards regarding the procurement, storage, dispensing, and supervision of hazardous chemicals. These measures have fortified the foundation of laboratory safety management. However, in practice, some research institutes simply replicate upper-level management systems without tailoring them to their specific circumstances, leading to inappropriate safety management methods for hazardous chemicals. In some cases, the established laboratory hazardous chemical safety management methods are merely to comply with various safety "supervision inspections," and the absence of dedicated management personnel leads to loopholes in the safety management of hazardous chemicals, diminishing their practical effectiveness in safety supervision.

Although university asset management and laboratory safety management departments conduct periodic inspections of the inventory and placement of regulated chemicals, difficulties in data acquisition and timeliness result in only obtaining "static" data at the time of inspection. This limits the capability for real-time monitoring of regulated chemical usage, making it challenging to rapidly trace chemical use. Consequently, issues such as idle chemicals and expired reagents not being safely disposed of promptly, difficulty in obtaining real-time ledgers, and discrepancies between records and actual inventory are common. These issues create safety hazards in every aspect of regulated chemicals' journey into the laboratory, significantly increasing laboratory risk factors.

2.2. Non-standardized storage and access

Universities involve a wide range of disciplines using regulated chemicals, covering hundreds of laboratories with highly dispersed storage, many of which involve regulated chemicals. The diverse types and characteristics of these chemicals significantly complicate the management of regulated chemicals, making uniform, standardized, and precise management challenging to achieve. Additionally, since the management of regulated chemicals involves multiple campus authorities and secondary units, the existing uni-directional vertical management model struggles to ensure timely and effective management processes and impedes data sharing among various management and user departments. This necessitates the adoption of target-oriented informational tools to assist in enhancing management efficiency and meeting the current research needs of universities.

University laboratories are characterized by a diversity of specialties, broad research directions, and rapidly updating research content. Consequently, the use of regulated chemicals also presents traits such as variety,

quantity, widespread distribution, and frequent procurement. Many universities still have considerable room for improvement in their management approaches to regulated chemicals. For instance, although a registration system is in place for the issuance and use of regulated chemicals, the variety and high frequency of use lead to frequent inventory movements, making registration work cumbersome. Coupled with the relatively small total volume used, the importance of these processes is often overlooked by users, leading to oversights. Manual registration methods not only consume significant manpower and resources and are inefficient, but they also fail to provide real-time insight into the procurement, storage, usage, and disposal of regulated chemicals. This can lead to inventory backlog, spoilage of chemicals, and untimely disposal of waste, increasing the safety management risks of regulated chemicals.

2.3. Weak awareness of hazardous chemical safety among scientific personnel

There is a cognitive blind spot among staff regarding the hazards and adverse effects of hazardous chemicals due to insufficient awareness. Some managerial department leaders are lax in their approach, leading to a diluted sense of safety management among staff members. A portion of the personnel also lack proper professional ethics, believing that they are not responsible for any safety incidents, thus neglecting management duties. Some staff members lack higher-level consciousness, do not proactively learn new laws, regulations, and management models, and have limited knowledge of hazardous chemicals, failing to manage them scientifically in categorized and batched manners. Moreover, they show indifference towards the transportation and usage status of hazardous chemicals. The absence of an assessment mechanism related to hazardous chemicals leads to a lack of focus in staff work. Some personnel perceive overly stringent management practices as impediments to the normal use of chemicals, thus harboring resistance to enhancing management levels.

2.4. Lack of health protection awareness among employees

In dealing with hazardous chemicals, staff members fail to comply with safety regulations by not wearing protective face shields, safety goggles, protective gloves, masks, and other safety equipment. Some leaders neglect the health issues of their staff, not conducting regular physical examinations. Additionally, the volatile nature of certain hazardous chemicals is overlooked, allowing their spread throughout corridors and laboratories, and even through air circulation systems into office areas, thereby affecting the health of the staff. When transporting hazardous gas cylinders, there is a lack of adherence to strict transportation standards, including the absence of shock-proof measures and uninstalled control valves. Moreover, there is a practice of touching communal items with gloves and protective clothing contaminated with hazardous chemicals [3].

3. Several recommendations for chemical management in China

3.1. Improving hazardous chemical safety management systems

The management of hazardous chemicals should implement the principle of ‘people-oriented, safety first, prevention-focused, and comprehensive management,’ aiming to create a safe, efficient, orderly, and clean laboratory working environment. Research institutions should earnestly implement national laws and regulations and establish a set of hazardous chemical safety management systems that align with their specific circumstances. These systems could include procurement, storage, dispensation, waste liquid collection, and responsibilities of hazardous chemical management personnel, to better ensure laboratory safety and efficient operation. In the implementation process, consideration could be given to establishing a laboratory management leadership team, led by the institute director, who would be fully responsible for the leadership and decision-making in hazardous chemical safety management. The deputy director in charge of safety production could

serve as the deputy team leader, responsible for the development and revision of laboratory regulations, supervision of hazardous chemical use safety, hidden danger investigation and supervision, and the disposal of laboratory waste and toxic substances. The team members would include laboratory management personnel, hazardous chemical management personnel, and personnel directly using hazardous chemicals, responsible for the daily storage, use, and custody of hazardous chemicals, as well as the collection and disposal of waste. After determining the responsible persons at all levels, a safety responsibility agreement should be signed to clarify their responsibilities, linking laboratory safety management issues with the performance of related personnel, and fundamentally addressing the issue of laboratory safety responsibility.

3.2. Digital recording of chemical use and storage

Radio Frequency Identification (RFID) technology, commonly abbreviated as RFID, is a method for using radio frequency communication to automatically identify and exchange information about non-contact moving or movable objects. Owing to its advantages such as its non-contact nature, large data capacity, high information processing speed, long recognition distance, and low cost, it is widely applied in fields like logistics, retail, manufacturing, apparel, and healthcare. For instance, RFID tags are used to collect data throughout the entire process from raw material procurement, production, and processing to warehouse distribution and final sales, enabling real-time monitoring and tracking of pharmaceuticals. International logistics giants like UPS, DHL, and FedEx are actively experimenting with RFID technology for widespread use in automatic information collection, cargo tracking, and warehouse management processes.

As a novel technology in hazardous chemical management, RFID is still in its infancy and faces many challenges during its development. However, the growth of technology is a process of twists and turns. The application of RFID in the production, storage, and transport management of hazardous chemicals can address issues of low production efficiency encountered in production. At the same time, it enhances supply chain management, improving its scientific nature. Furthermore, it can fill the gaps in the logistics management of hazardous chemicals, ensuring their safety and quality to the greatest extent. With the development of RFID technology, a revolution in the field of hazardous chemical logistics management is emerging, which will become a growth point in the future economy, creating significant value [4].

Traditional laboratory hazardous chemical labels are prone to falling off and slow to update information. The use of RFID electronic tags brings significant innovation to traditional hazardous chemical storage methods. Consisting of a coil, control circuit, and storage device, RFID electronic tags store identifiable data about objects, such as numbers, names, purchase times, usage records, and responsible persons, and can be integrated with intelligent control cabinets for real-time data updates and maintenance. As the unique identification mark of each inventory hazardous chemical, RFID tags are fundamental to the entire inventory management of hazardous chemicals. Scanning RFID tags with an RFID handheld terminal forms a dynamic electronic ledger from stock entry to waste bottle disposal, which is key to establishing a sound, scientific, efficient, and standardized hazardous chemical safety management system [5].

Leveraging information technology to establish a ‘one-code-per-bottle’ system for controlled chemicals is crucial. Each bottle of reagent is assigned a unique code (compatible with RFID technology and QR code scanning) before storage, utilizing RFID technology for inventory management, accurately recording electronic information of chemicals, and ensuring real-time, dynamic management of controlled chemicals. Through various modules such as laboratory management (safety education exams, electronic usage ledgers), procurement management (unified management, standardized procurement processes), storage management (stock shortage alerts, scan-to-stock, environmental monitoring), and inspection management (user information

authorization verification, reagent usage records, safety status, reagent inventory), controlled chemicals can be monitored informatively at each stage, realizing real-time supervision of the entire lifecycle from the application, approval, storage, placement, inquiry, dispensation, recycling, to disposal, ensuring online data, network collaboration, traceability, and encrypted data storage, achieving integrated management of 'people, machine, materials, and environment.'

3.3. Strengthening safety training for laboratory personnel

Inviting professionals to conduct lectures or promotional activities in laboratories, or using brochures or videos to strengthen staff awareness of relevant laws, regulations, and laboratory systems, and to enhance their safety consciousness. Training on responsibility awareness should also be strengthened, promoting staff awareness of responsibility, so that they act conscientiously and responsibly during work, avoiding oversights in the management of hazardous chemicals due to human factors.

Periodic training using hazardous chemical safety accident case studies as warnings should be conducted. Various forms of warning education, such as accident videos, promotional boards, and firsthand accounts from victims, can be used to enhance staff safety awareness and serve as a deterrent.

Regular and appropriate emergency drills for hazardous chemical accidents should be conducted, such as how to respond to inhalation of hazardous chemical gases, emergency treatment after contact with hazardous chemical reagents, and how to respond to laboratory fires or explosions. Emergency drills can improve staff response speed in emergencies and enhance the emergency handling capabilities of all members [3].

4. Conclusion

In conclusion, the effective management of hazardous chemicals in Chinese research institutions is a multifaceted challenge that demands comprehensive and meticulous approaches. The current scenario, characterized by gaps in safety management systems, insufficient awareness among scientific personnel, and outdated storage and tracking methods, underscores the need for systemic reform and innovation.

To address these challenges, it is imperative to refine and implement safety management systems that are tailored to the specific needs of each institution, grounded in the principles of prioritizing human safety and employing a prevention-focused strategy. The integration of digital technologies, such as RFID tagging, plays a pivotal role in transforming the management of hazardous chemicals from a traditional, static approach to a dynamic, real-time, and efficient system. This technological shift not only enhances the accuracy of chemical tracking but also ensures compliance with safety regulations and facilitates a responsive management process. Furthermore, the reinforcement of safety consciousness among laboratory personnel is crucial. Through regular training, emergency drills, and the dissemination of information on hazardous chemical management and health protection, a culture of safety and responsibility can be cultivated among staff and management alike. Such an approach not only mitigates the risk of accidents but also promotes a proactive stance towards safety and health concerns.

In light of these recommendations, it is clear that addressing the current deficiencies in hazardous chemical management requires a holistic approach. This involves not only upgrading management systems and adopting new technologies but also fostering a culture of safety and responsibility at all levels of the institution. By doing so, research institutions in China can create a safer, more efficient, and more conducive environment for scientific advancement, ultimately contributing to the broader goal of societal progress and stability.

Disclosure statement

The author declares no conflict of interest.

References

[1] Meng Z, Yan X, Liu B, et al., 2023, Exploration and Practice of a Smart Management Solution for Controlled Chemicals in Universities. Laboratory Research and Exploration, 2023(42): 275–280.

[2] Tang M, Wen J, Cheng S, et al., 2023, Analysis of Safety Issues and Strategies for Hazardous Chemicals in Research Institutes. Chemical Engineering Management, 2023(27): 92–94.

[3] Zhang L, 2023, Discussion on Laboratory Hazardous Chemical Management. China Petroleum and Chemical Standards and Quality, 43(18): 68–70.

[4] Wang G, 2022, Application of RFID Technology in the Management of Hazardous Chemicals. Chemical Engineering Design Communications, 48(12): 150–152.

[5] Cao C, Wang C, Yu C, et al., 2023, Exploration of Laboratory Hazardous Chemical Management Based on RFID Technology. Journal of Zhejiang Wanli University, 36(6): 82–87.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Research on Modulation Signal Denoising Method Based on Improved Variational Mode Decomposition

Canyu Mo[1,2], Qianqiang Lin[2]*, Yuanduo Niu[1,2], Haoran Du[1,2]

[1]Xi'an Electronic Engineering Research Institute, Xi'an 710075, China
[2]College of Electronic Science, National University of Defense Technology, Changsha 410199, China

***Corresponding author:** Qianqiang Lin, even_qqlin@nudt.edu.cn

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In order to further analyze the micro-motion modulation signals generated by rotating components and extract micro-motion features, a modulation signal denoising algorithm based on improved variational mode decomposition (VMD) is proposed. To improve the time-frequency performance, this method decomposes the data into narrowband signals and analyzes the internal energy and frequency variations within the signal. Genetic algorithms are used to adaptively optimize the mode number and bandwidth control parameters in the process of VMD. This approach aims to obtain the optimal parameter combination and perform mode decomposition on the micro-motion modulation signal. The optimal mode number and quadratic penalty factor for VMD are determined. Based on the optimal values of the mode number and quadratic penalty factor, the original signal is decomposed using VMD, resulting in optimal mode number intrinsic mode function (IMF) components. The effective modes are then reconstructed with the denoised modes, achieving signal denoising. Through experimental data verification, the proposed algorithm demonstrates effective denoising of modulation signals. In simulation data validation, the algorithm achieves the highest signal-to-noise ratio (SNR) and exhibits the best performance.

Keywords: Micro-motion modulation signal; Variational mode decomposition; Genetic algorithm; Adaptive optimization

Online publication: January 18, 2024

1. Introduction

With the continuous development of detection technology, the collection environment with actual rotating component targets has become increasingly complex. Due to the presence of ground clutter and interfering noise, the collected modulated echo data is often mixed with various noise interferences, making the effective modulation information unclear and resulting in a low signal-to-noise ratio (SNR) [1-3]. In order to achieve the goal of high SNR for micro-motion modulated echoes, exploring and researching new denoising methods has become an inevitable trend. As a result, numerous domestic and foreign scholars have conducted research on methods to improve the SNR of radar echoes. Among them, the method based on signal decomposition and

reconstruction has achieved good results in clutter suppression and denoising of Doppler signals collected from targets with rotating components [4-6].

Fourier transform denoising is a method of global transformation that lacks local descriptive ability. During denoising, it will lose a large amount of effective signal, resulting in poor processing results for non-stationary signals [7]. Wavelet threshold denoising transforms micro-motion modulated signals into the time-scale domain, but it requires the selection of appropriate wavelet functions and thresholds to achieve a better denoising effect [8]. However, these methods are all based on fixed transform basis functions and cannot adaptively process structurally complex modulated echo data. The Empirical Mode Decomposition (EMD) [9] method based on signal decomposition theory can adaptively decompose the signal itself, dividing it into a series of Intrinsic Mode Functions (IMFs) that represent local energy anomalies within different frequency bands. However, this method is prone to mode mixing, boundary effects, and other issues. The Ensemble EMD (EEMD) and Complete Ensemble EMD (CEEMD) methods are further improvements upon the EMD method. EMD, EEMD, and CEEMD are all recursive signal processing methods, but they have limitations when dealing with multi-component modulated signals [10]. The non-recursive signal decomposition method, Variational Mode Decomposition (VMD) eliminates the process of selecting components and transfers the acquisition process of signal intrinsic mode function (IMF) components to a variational framework [11]. This allows for adaptive and effective separation of the frequency domain portion of the signal and its components. Compared to methods like EMD, VMD has a solid theoretical foundation and effectively avoids mode mixing problems. It also exhibits better noise robustness. Although the VMD method effectively decomposes the signal into several sub-signals with different center frequencies, several parameter values must be set before the decomposition to achieve the best results. Typically, manually selecting parameters cannot produce optimal results. Sarangi and colleagues utilized a particle swarm optimization algorithm to simultaneously optimize the values of K and α for an optimal solution [12]. Naik and colleagues used the newly proposed orthogonal low sidelobe as an optimization objective and applied the firefly algorithm to search for optimal parameter combinations (K,α) [13].

Based on this, this paper combines the genetic algorithm with the VMD algorithm and applies it to denoise micro-motion modulated echoes. This method effectively extracts modulation information while adaptively selecting the determining parameters of VMD, minimizing signal loss as much as possible. Both theoretical models and practical verification results demonstrate that the application of the improved method proposed in this paper exhibits good performance in denoising, effectively improving the signal-to-noise ratio of micro-motion modulated signals. This lays a solid foundation for further analysis and processing of modulation components.

2. GA-VMD denoising method

2.1. Variational mode decomposition

The goal of VMD is to decompose the original signal into intrinsic mode functions (IMFs) with fixed bandwidth and center frequencies. The decomposition process of the original signal is essentially a solving process of a variational problem.

The expression for the intrinsic mode functions is as follows:

$$u_k(t) = A_k(t)\cos\left(\Phi_k(t)\right) \qquad (1)$$

where $A_k(t)$ represents the instantaneous amplitude of $u_k(t)$. $\Phi_k(t)$ represents the phase, which is non-monotonically decreasing.

If the demodulation method is used to estimate the bandwidth of IMF, it will lead to a variational constrained problem. The expression for the variational constrained model is as follows:

$$min_{\{u_k\},\{w_k\}}\left\{\sum_k \left\|\partial_i\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-jw_k t}\right\|_2^2\right\} \quad s.t.\ \sum_k u_k = f \tag{2}$$

where $\{u_k\}=\{u_1,\ldots,u_k\}$ and $\{W_k\}=\{W_1,\ldots,W_k\}$ represent the IMFs and their corresponding center frequencies, respectively. K represents the number of decomposition layers.

To obtain the optimal solution of this variational constrained model, the algorithm introduces the Lagrange multiplier λ operator and quadratic penalty function term α. The expression of the augmented Lagrange function is as follows:

$$L(\{u_k\},\{w_k\},\lambda) = \alpha\sum_k \left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-jw_k t}\right\|_2^2 + \left\|f(t)-\sum_k u_k(t)\right\|_2^2 + \left\langle\lambda(t), f(t)-\sum_k u_k(t)\right\rangle \tag{3}$$

Finally, the variational constrained model is solved using the Alternating Direction Method of Multipliers (ADMM). The expressions for the updated formulas of $\{u_k\}$, $\{W_k\}$, and λ are as follows:

$$u_k^{n+1}(W) = \frac{f(W)-\sum_{i+k}\hat{u}_i(W)+\frac{\lambda^n(W)}{2}}{1+2\alpha(w-w_k)^2} \tag{4}$$

$$w_k^{n+1} = \frac{\int_0^\infty w|\hat{u}_k(W)|^2 dw}{\int_0^n |\hat{u}_k(W)|^2 dw} \tag{5}$$

$$\lambda^{n+1}(W) - \lambda^n(W) + \tau\left[f(W)-\sum_k u_k^{n+1}(W)\right] \tag{6}$$

The termination constraint condition is as follows:

$$\sum_k \left\|u_k^{n+1} - \frac{u_k^n\|_2^2}{\|u_k^n\|_2^2} < \varepsilon \right. \tag{7}$$

where ε represents the convergence accuracy.

2.2. Genetic algorithm-based parameter optimization for VMD

In the VMD algorithm, the penalty parameter and the number of mode components K need to be preset based on prior experience. When the value of K is too large, it can lead to excessive decomposition of the signal. On the other hand, when K is too small, it becomes difficult to effectively separate the correct center frequencies of the modes, resulting in mode mixing phenomena. Therefore, selecting the appropriate parameter combination is

crucial for denoising using VMD. Here, a genetic algorithm is employed to optimize the values of parameters K and α, ensuring that each detection of the micro-vibration modulation signal is optimal and effectively reducing the randomness in parameter selection.

The genetic algorithm is an adaptive optimization algorithm with excellent global probability search capability. The process of optimizing parameters mainly includes steps such as initial coding, generating initial population, setting fitness parameters, selection, crossover, and mutation.

(1) The punishment parameter α and the number of mode components K are encoded and initialized for chromosome initialization, generating an initial population that contains different combinations of (K,α) individuals.

(2) The signal is decomposed using different combinations within each individual, and the multiscale sample entropy of each decomposed mode function is calculated to determine the corresponding fitness. The local minimum is defined as the minimum value of the sample entropy.

(3) Using the fitness values obtained in step (2), selection is performed on the superior individuals in the initial population, followed by crossover and mutation to form a new generation population. This process is iterated continuously, comparing the local fitness values to find the global minimum.

(4) Stop the iteration and determine the optimal solution for the parameter combination (K,α) .

2.3. Combining genetic algorithm with VMD for denoising

Given the nonlinear and non-stationary characteristics of micro-vibration modulation signals, as well as the suboptimal performance of a single VMD denoising method, this paper proposes a VMD decomposition approach for denoising modulated noisy signals. By optimizing the decomposition parameters using the GA algorithm, the reconstructed modulated signal is obtained. The following is the specific implementation process of the proposed method in this paper.

(1) Using genetic algorithm to optimize VMD parameters, enabling accurate decomposition of the original signal.

(2) Using the VMD method to decompose the acquired modulated signal.

(3) Reconstructing the original signal to obtain the denoised result.

To verify the effectiveness of the proposed algorithm, two objective performance indicators, SNR and root mean square error (RMSE), are used to evaluate the denoising performance of various methods. The SNR parameter reflects the denoising ability of the method, with a larger SNR indicating better denoising performance. The RMSE reflects the difference in the signal amplitude before and after denoising, with a smaller RMSE indicating better denoising performance.

$$SNR = 10lg\frac{\sum_{i=1}^{n} f(t_i)^2}{\sum_{i=1}^{n}\left[\hat{f}(t_i) - f(t_i)\right]^2} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left[\hat{f}(t_i) - f(t_i)\right]^2}{n}} \quad (9)$$

3. Experimental data verification

The performance of the denoising method is evaluated using two objective parameters, SNR and RMSE. Gaussian white noise with variances ranging from 0.01 to 0.09 is added to the simulated helicopter echo data

with a diameter of 5.029 m and a rotational speed of 6.8 r/s.

Table 1. Comparison of SNR values

Noise variance	VMD	GA-VMD
0.01	15.127	15.489
0.02	14.527	15.153
0.03	14.425	14.857
0.04	14.127	14.852
0.05	13.256	13.527
0.06	12.775	13.578
0.07	12.084	12.578
0.08	10.852	11.575
0.09	10.052	11.425

Table 1 lists the comparison of values between the traditional VMD denoising method and the proposed method in this paper for noise variances ranging from 0.01 to 0.09. From the table, it can be observed that the GA-VMD method proposed in this paper has the lowest SNR value. Moreover, because VMD itself is a powerful tool for analyzing non-stationary and nonlinear signals, it has a solid mathematical theoretical foundation. As long as the parameters are properly selected, the issue of mode mixing can be avoided. Therefore, in this paper, a genetic algorithm is used to optimize the two parameters that affect the decomposition performance, achieving accurate signal decomposition. The VMD method operates in the frequency domain, which also results in relatively high computational efficiency. The decomposed components can be considered as stable signals. By reconstructing the original signal, the denoising effect is achieved. When the noise variance ranges from 0.01 to 0.09, the proposed method achieves a maximum SNR value. Compared to the traditional VMD denoising method, it is higher by 0.271 to 0.803. As the noise intensity increases, this difference also increases.

Table 2. Comparison of RMSE values

Noise variance	VMD	GA-VMD
0.01	0.0387	0.0325
0.02	0.0395	0.0311
0.03	0.0414	0.0411
0.04	0.0426	0.0419
0.05	0.0448	0.0438
0.06	0.0458	0.0442
0.07	0.0462	0.0458
0.08	0.0483	0.0467
0.09	0.0502	0.0481

Table 2 presents the RMSE values of the two methods when the noise variance ranges from 0.01 to 0.09. From the table, it can be observed that at the same noise variance, the proposed method achieves the lowest

RMSE value. Compared to the traditional VMD denoising method, the RMSE value of the proposed method is reduced by 0.004 to 0.0021. As the noise intensity increases, this difference also increases.

To validate the effectiveness of the proposed algorithm for denoising real measurement data, the Robinson R44 helicopter is taken as an example. The target is hovering under the far-field conditions of an S-band radar. At this time, the azimuth angle α=149° and the elevation angle β=9.15°. The specific radar parameters are shown in **Table 3**.

Table 3. Main parameters of the linear frequency modulation (LFM) pulse signal

Parameters	Values	Unit
Radio frequency	2.94912	GHz
Bandwidth	120	MHz
Pulse width	2.0835	**μs**
Pulse repetition frequency	3.7202	kHz
Number of range samples	8,192	piece
Slow time dimension	4,096	piece
Sampling frequency	491.52	MHz

The specific parameters of the rotating components of the Robinson R44 helicopter are shown in **Table 4**.

Table 4. Structural parameters of the R44 helicopter

Parameters	Values	Unit
Number of rotor blades	1	piece
Number of blades on a single rotor	2	piece
Blade length	5.029	m
Blade rotational speed	6.8	rad/s

Performing a short-time Fourier transform directly on the modulated component echoes yields their spectrogram, as shown in **Figure 1(a)**. From the figure, it can be seen that due to the presence of background noise and clutter interference, the signal-to-noise ratio of the modulated component's spectrogram is not high, and the noise interference is mixed in the echo signal. To eliminate these effects and better reflect the characteristics of the micro-motion modulated signal, traditional VMD and the method proposed in this paper are used for denoising and comparison. **Figures 1(b)** and **(c)** respectively show the modulated signals obtained after denoising using the traditional VMD method. The traditional VMD method, due to its lack of adaptability in the selection of , often resulting in either over-decomposition or under-decomposition of the signal, leading to less-than-ideal denoising results.

From **Figure 1**, it can be seen that the method proposed in this paper achieves more thorough noise removal. The processed in-phase axis is clearer, and some information that was masked by high-frequency noise is effectively recovered. The denoising effect is relatively ideal. To further demonstrate the superiority of the method proposed in this paper, 1,841 distance units on the flight route were selected for analysis. **Figure 2** shows the optimization curves of the Genetic Algorithm-VMD for each of these distance units.

From **Figure 2**, it can be observed that the fitness value of the data becomes stable when the genetic algorithm reaches generation 6, indicating the best decomposition effect. Moreover, compared to other methods,

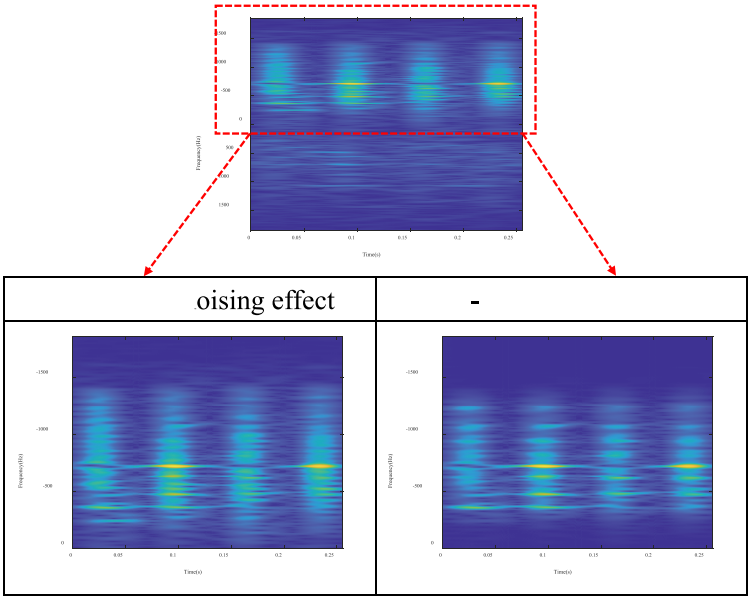


Figure 1. Comparison of measured denoising results

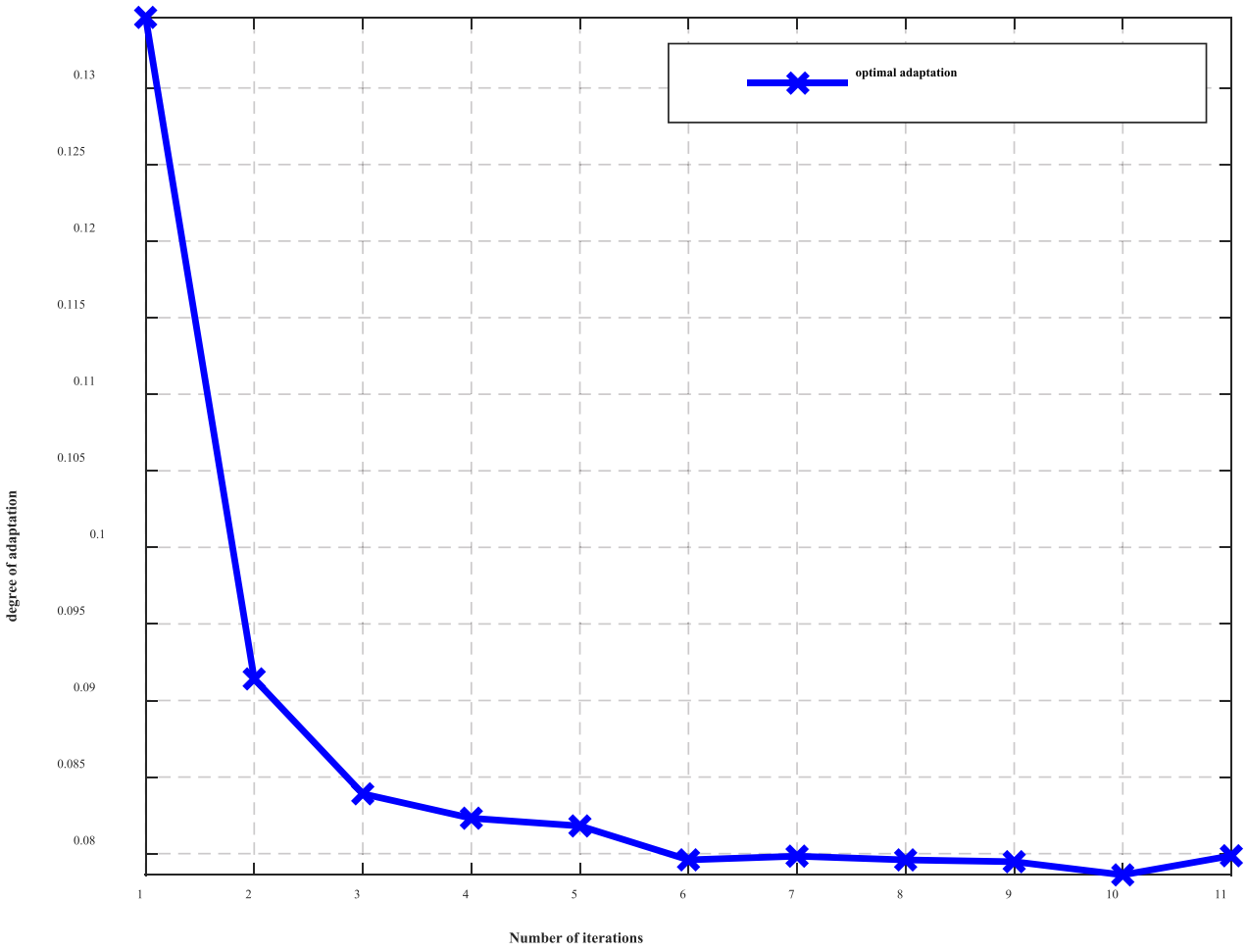


Figure 2. Optimization curves

the proposed method in this paper preserves a wider frequency band while effectively suppressing high-frequency noise, resulting in a relatively ideal denoising effect.

4. Conclusion

The proposed method in this paper adopts a combination of VMD and genetic algorithm to optimize the decomposition parameters. This approach leverages the adaptability of VMD decomposition, the strong mathematical theoretical foundation and the high-frequency noise suppression capability of the genetic algorithm itself. The effectiveness of the proposed method in practical application scenarios is validated through two objective parameters: SNR and root mean square error (RMSE). The experimental results indicate that the proposed method in this paper achieves a significant improvement in signal-to-noise ratio compared to other methods. Furthermore, as the noise intensity increases, the denoising effect becomes more pronounced.

The drawback of the proposed method in this paper is that the VMD method requires a lot of parameter settings, and it is necessary to select appropriate parameters for accurate signal decomposition, which, in turn, leads to better decomposition results. Although the genetic algorithm used in this paper is an intelligent optimization algorithm, it does have a significant computational complexity in terms of optimization.

Disclosure statement

The authors declare no conflict of interest.

Author contributions

Conceptualization: Canyu Mo
Formal analysis: Yuanduo Niu
Writing – original draft: Canyu Mo
Writing – review & editing: Qianqiang Lin, Haoran Du

References

[1] Zhu L, Zhang S, Chen K, et al., 2022, Low-SNR Recognition of UAV-to-Ground Targets Based on Micro-Doppler Signatures Using Deep Convolutional Denoising Encoders and Deep Residual Learning. IEEE Transactions on Geoscience and Remote Sensing, 60: 1–13. https://doi.org/10.1109/TGRS.2021.3123109

[2] Xu X, Feng C, He S, 2020, A Method for the Micro-Motion Signal Separation and Micro-Doppler Extraction for the Space Precession Target. IEEE Access, 8: 130392–130404. https://doi.org/10.1109/ACCESS.2020.3008480

[3] Yang Y, Wen P, Ye W, et al., 2023, Blind Universal Denoising for Radar Micro-Doppler Spectrograms Using Identical Dual Learning and Reciprocal Adversarial Training. IEEE Transactions on Circuits and Systems for Video Technology. https://doi.org/10.1109/TCSVT.2023.3323985

[4] Ram SS, Vishwakarma S, Sneh A, et al., 2021, Sparsity-Based Autoencoders for Denoising Cluttered Radar Signatures. IET Radar, Sonar & Navigation, 15(8): 915–931. https://doi.org/10.1049/rsn2.12065

[5] Ding Y, Tang J, 2014, Micro-Doppler Trajectory Estimation of Pedestrians Using a Continuous-Wave Radar. IEEE Transactions on Geoscience and Remote Sensing, 52(9): 5807–5819. https://doi.org/10.1109/TGRS.2013.2292826

[6] Du L, Wang B, Wang P, et al., 2015, Noise Reduction Method Based on Principal Component Analysis with Beta Process for Micro-Doppler Radar Signatures. IEEE Journal of Selected Topics in Applied Earth Observations and

Remote Sensing, 8(8): 4028–4040. https://doi.org/10.1109/JSTARS.2015.2451004

[7] Song D, Chung Baek AM, Kim N, 2021, Forecasting Stock Market Indices Using Padding-Based Fourier Transform Denoising and Time Series Deep Learning Models. IEEE Access, 9: 83786–83796. https://doi.org/10.1109/ACCESS.2021.3086537

[8] Ouyang C, Cai L, Liu B, et al., 2023, An Improved Wavelet Threshold Denoising Approach for Surface Electromyography Signal. EURASIP Journal on Advances in Signal Processing, 2023: 108. https://doi.org/10.1186/s13634-023-01066-3

[9] Young H-WV, Lin Y-C, Wang Y-H, 2022, On the Memory Cost of EMD Algorithm. IEEE Access, 10: 114242–114251. https://doi.org/10.1109/ACCESS.2022.3218417

[10] Li W, Xu H, Jiang B, et al., 2023, Wavelet Threshold Ultrasound Echo Signal Denoising Algorithm Based on CEEMDAN. Electronics, 12(14): 3026. https://doi.org/10.3390/electronics12143026

[11] Chen G, Zhang T, Qu W, et al., 2023, Photovoltaic Power Prediction Based on VMD-BRNN-TSP. Mathematics, 11(4): 1033. https://doi.org/10.3390/math11041033

[12] Sarangi S, Dash PK, Bisoi R, 2023, Probabilistic Prediction of Winder Speed Using an Integrated Deep Belief Network Optimized by a Hybrid Multi-Objective Particle Swarm Algorithm. Engineering Applications of Artificial Intelligence, 126(PC): 107034. https://doi.org/10.1016/j.engappai.2023.107034

[13] Naik J, Bisoi R, Dash PK, 2018, Prediction Interval Forecasting of Wind Speed and Wind Power Using Modes Decomposition Based Low Rank Multi-Kernel Ridge Regression. Renewable Energy, 129(PA): 357–383. https://doi.org/10.1016/j.renene.2018.05.031

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Instrument Automation Control System Faults and Maintenance

Jianwei Li*

China Institute for Radiation Protection, Taiyuan 030006, Shanxi Province, China

***Corresponding author:** Jianwei Li, 277102630@qq.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: This article explores the topic of fault diagnosis and maintenance strategies for instrument automation control systems, analyzing them through specific cases. The aim of this research is to improve the stability and reliability of the system by conducting a thorough investigation of faults and maintenance in instrument automation control systems. By doing so, this research hopes to provide a strong guarantee for the smooth progress of industrial production.

Keywords: Instrument automation; Control system failure; Maintenance

Online publication: January 18, 2024

1. Preface

Industrial production automation represents an important trend in the development of modern industry. It not only improves production efficiency and quality but also reduces labor costs and human errors. The instrument automation control system constitutes a vital component of industrial production automation, with its operational status significantly impacting the stability and reliability of the entire industrial production process.

The technology and hardware of instrument automation control systems are continually evolving. New technologies and hardware upgrades contribute to improved system performance and stability as well as enhance the maintainability and scalability of the system. The instrument automation control system is playing an increasingly pivotal role in industrial production. However, owing to factors such as the complexity of the control system and prolonged high-intensity operation, various faults may arise in the instrument automation control system, including abnormal pressure, excessive temperature, material leakage, and more. These safety hazards pose a significant risk of causing serious harm to personnel and equipment and may even lead to major safety accidents [1-6]. Therefore, conducting in-depth research on the faults and maintenance of instrument automation control systems holds immense practical significance.

2. Instrument automation control system failure

2.1. Control system hardware failure

(1) Power failure: This may be caused by insufficient power supply, poor contact, or a short circuit in the power line. A power failure can lead to improper functioning of the control system.

(2) Controller failure: The controller, a core component of the instrument automation control system, can cause the entire system to malfunction if it fails [7]. Controller failure may result from hardware or software issues.

(3) Sensor failure: Sensors, crucial measuring devices in the instrument automation control system, can compromise measurement accuracy when they fail. Sensor failure may stem from aging, wear, poor contact, or a short circuit in the connecting wire.

(4) Actuator failure: Actuators, vital output devices of the control system, can disrupt normal execution if they fail [8]. Actuator failure may result from hardware or software issues.

2.2. Control system software failure

(1) Control program errors: Software failures may be caused by errors in the control program, stemming from design flaws, incorrect programming language, or logic errors.

(2) Human-machine interface failure: The human-machine interface, the link between the control system and the operator, can impede normal operation if it fails. Failures may include a non-responsive display screen, malfunctioning buttons, an insensitive touch screen, or errors in serial port communication.

(3) Data processing errors: Data processing, a crucial function of the control system, may lead to abnormal operation if errors occur [9]. Issues may arise from incorrect data input, flawed data processing algorithms, or inaccurate data output.

2.3. Failures caused by external factors

(1) Environmental factors: High temperature, low temperature, moisture, dust, and vibration are environmental factors that may adversely affect the hardware and software of the control system, leading to failure.

(2) Lightning and surge: Common natural disasters like lightning and surges can damage the control system hardware and software. Protective measures should be taken to prevent such damage.

(3) Electromagnetic interference: Electromagnetic interference may adversely affect signal transmission and data processing, causing failure [10]. To mitigate this impact, appropriate electromagnetic shielding and filtering measures are necessary.

(4) Power supply system failure: Issues such as an unstable power supply, poor contact, or a short circuit in the power line may affect the normal operation of the control system, leading to hardware and software damage. Ensuring the stability and reliability of the power supply system requires protective measures, such as equipping backup power supplies or using uninterruptible power supplies.

3. Instrument automation control system fault maintenance methods

3.1. Troubleshooting

(1) Direct observation method: Observe the instrument's appearance, indicator lights, and display panel to check for abnormalities such as damage, non-illuminated indicator lights, or abnormal displays [11].

(2) Auditory method: Listen carefully to the equipment's running sounds, such as the rotation of the fan or alarm sounds [12]. Unusual sounds may indicate a device malfunction.
(3) Touch method: Physically touch the device's surface and circuit connections to detect overheating, looseness, etc.
(4) Smell method: If there are odors such as burning or burnt wires, the equipment may be malfunctioning.
(5) Instrument detection method: Utilize multimeters, oscilloscopes, and other tools to measure voltage, current, and other parameters, determining the fault location and cause.

3.2. Daily maintenance

Regularly inspect the instrument automation control system, including hardware, software, power supply systems, and other components. Pay attention to the equipment's operating status, line connections, cooling device conditions, etc. Regularly clean and maintain the control system equipment, keeping the equipment's surface clean and preventing dust and debris from affecting its normal operation. Perform regular backups of crucial control system data, including system settings, control programs, historical data, etc., to prevent data loss or damage [13-15].

3.3. Troubleshooting

Based on the fault manifestation, determine the type and possible causes, such as hardware equipment failure, software system failure, power supply system failure, etc. [16]. Formulate corresponding maintenance plans based on the fault type. For hardware equipment failure, consider replacing spare parts or repairing damaged equipment. For software system failure, reinstalling or updating the software may be necessary. For power supply system failure, check the power line or equip a backup power supply. During maintenance, prioritize safe operations and protective measures. Ensure the quality and compatibility of spare parts if replacements or repairs are needed. Ensure the stability and compatibility of the software during reinstallation or updates [17]. After maintenance, test the control system to ensure its normal operation. Record and summarize the maintenance process for future reference.

3.4. Preventive measures

Regularly perform preventive maintenance, including replacing worn parts, tightening loose screws, cleaning heat dissipation devices, etc., to prevent equipment failures. Establish maintenance files for the control system, recording equipment maintenance history and spare parts replacement, facilitating comprehensive management and maintenance [18-20]. Conduct training and technical upgrades for maintenance personnel to improve their skill level and professionalism in maintaining the control system. Strengthen security protection to prevent unauthorized access and malicious attacks, ensuring the security and stability of the system.

4. Case analysis

4.1. Case 1

A petrochemical enterprise boasts an advanced instrument automation control system primarily utilized for monitoring and controlling the production process. However, recent operational issues have led to a decrease in production efficiency. To swiftly restore normal system operation, the company decided to conduct fault diagnosis and repair of the control system.

4.1.1. Troubleshooting

Firstly, examine the hardware equipment of the control system, including sensors, actuators, controllers, etc. The inspection revealed that the signal transmission lines of some sensors were aging, causing unstable signal transmission. Secondly, test the software functions of the control system. Anomalies were identified in some functional modules. After analysis, it was determined that software version incompatibility was the root cause. Next, scrutinize the network communication of the control system, revealing interference and signal attenuation problems leading to data transmission delays and losses.

4.1.2. Maintenance process

Based on the fault diagnosis results, prepare corresponding maintenance tools and spare parts, such as multimeters, oscilloscopes, spare sensors, and software upgrade packages. Replace seriously aging sensor lines to ensure stable signal transmission. Resolve issues with incompatible software versions by either upgrading the software or rolling back to a stable version. Optimize and adjust network communication lines, incorporating signal amplifiers and anti-interference equipment to address communication problems. Debug and calibrate the repaired system to ensure all functions operate normally. Record the fault diagnosis and maintenance process in detail to provide a reference for future maintenance work. After passing the company's acceptance, the control system is delivered back to the production department for use. Propose preventive measures for the cause of the fault and the maintenance process, such as regularly checking hardware equipment, upgrading software versions, optimizing network configuration, etc.

4.1.3. Result analysis

This fault diagnosis and maintenance process involves hardware inspection, software functional testing, network communication debugging, and other aspects. Through steps such as preparing maintenance tools and spare parts, replacing faulty parts, repairing software, and adjusting network communications, the problems in the control system were ultimately resolved. Simultaneously, preventive maintenance recommendations are presented to reduce the incidence of similar failures and enhance the stability and reliability of the control system.

4.2. Case 2

The DCS control system of a chemical plant suddenly crashed during the production process. Upon inspection, it was discovered that a key component of the operator station failed. Due to the timely replacement of spare parts, the system quickly returned to normal operation. This case underscores the importance of equipping spare parts for critical equipment and establishing an emergency response mechanism. When a failure occurs, measures should be promptly taken to restore the normal operation of the equipment.

4.3. Case 3

The automated control system of a sewage treatment plant often experiences data transmission delays and data loss. Upon investigation, it was identified that interference and signal attenuation problems existed in the network communication lines. Optimizing the network configuration and adding signal amplifiers resolved the issue. This case highlights the significance of focusing on the maintenance and optimization of the network communication component for systems involving data transmission. Targeted measures should be implemented to address problems promptly when they arise.

4.4. Case 4

During the production process, a steel plant's heating furnace automation control system suddenly failed to control the furnace temperature. Upon inspection, it was revealed that the insulation layer of the sensor circuit was damaged, causing signal interference. After repairing the circuit, the system returned to normal operation. This case underscores the necessity of regularly inspecting and maintaining sensor signal transmission lines. Promptly addressing any damage or aging in the lines is crucial to avoiding potential risks of failure.

5. Summary

This article emphasizes the importance of system maintenance by exploring fault diagnosis and maintenance strategies for instrument automation control systems and analyzing specific cases. In practical applications, it is essential to formulate corresponding maintenance plans and measures tailored to the system's characteristics and the operating environment. Regular inspections, preventive maintenance, emergency interventions, and technical upgrades are crucial avenues to effectively enhance the stability and reliability of the instrument automation control system. These efforts provide a robust guarantee for the seamless advancement of industrial production.

Disclosure statement

The author declares no conflict of interest.

References

[1] Wang J, Jia F, 2023, Failure and Maintenance of Electrical Instruments in Automatic Control Systems. Chemical Industry Management, 2023(20): 127–130.

[2] Liu X, 2023, Research on Instrument Automation Control System Faults and Maintenance. Modern Manufacturing Technology and Equipment, 59(6): 180–182.

[3] Chu C, He Y, Ji B, et al., 2022, Preventive Maintenance of Flow Measurement System. Petrochemical Automation, 58(6): 75–80.

[4] Huang L, Lin Y, 2022, Failure and Maintenance Analysis of Chemical Instrument Automatic Control System. China Equipment Engineering, 2022(S1): 16–18.

[5] Wang Z, 2022, Exploration and Practice of Preventive Maintenance Strategies for Instrument Control Systems in Large Refining and Chemical Enterprises. Petrochemical Automation, 58(4): 6–11.

[6] Cao Y, Ding D, Yu M, 2022, Analysis of Maintenance Technology of Automated Instrumentation and Control Systems. Integrated Circuit Applications, 39(7): 248–249.

[7] Xi S, 2022, Discussion on Instrument Automation Control System Faults and Maintenance Technology. Metallurgy and Materials, 42(3): 96–98.

[8] Sun H, 2022, Failure and Maintenance of Chemical Instrument Automatic Control System. Chemical Engineering and Equipment, 2022(2): 165–166.

[9] Fu X, 2021, Key Technologies for Analysis and Troubleshooting of Automobile Engine Electronic Control Systems. Internal Combustion Engines and Accessories, 2021(10): 160–161.

[10] Sun Y, 2021, A Nonlinear System Actuator Fault Estimation Method Containing Disturbance Terms. Fire Power and Command and Control, 46(3): 38–42.

[11] Zhou X, 2020, Failure and Maintenance Technology Analysis of Chemical Instrument Automatic Control System.

Electronic Test, 2020(22): 117–118 + 59.

[12] Guan X, Wei B, Meng X, 2019, Chemical Instrument Automatic Control System Failure and Maintenance Technology. Polyvinyl Chloride, 47(12): 19–21.

[13] Pan D, 2019, Analysis of Instrument Automation Control System Faults and Maintenance Technology. Petrochemical Technology, 26(9): 222–223.

[14] Liu T, Huang H, 2019, Discussion on Fault and Maintenance Technology of Instrument Automation Control System. China New Communications, 21(15): 148–149.

[15] Yang D, 2019, Faults and Maintenance of Instrument Automation Control System. Petrochemical Technology, 26(3): 217.

[16] Yang L, 2019, Analysis of Instrument Automation Control System Faults and Maintenance Technology. Electronic World, 2019(3): 198 + 200.

[17] Su Z, 2016, Discussion on Fault and Maintenance Technology of Instrument Automation Control System. China High-Tech Enterprises, 2016(12): 59–60.

[18] Lai Y, 2015, Analysis of Faults in On-Site Instrument Control Systems. Science and Technology Vision, 2015(23): 248.

[19] Liao X, Chen Y, 2015, Exploring the Fault Analysis and Countermeasures of Boiler Instrument Control System. Information and Communications, 2015(6): 91.

[20] Wang R, Wang H, Li G, 2014, Analysis of Common Faults of Production Process Control Instruments. China Chlor-Alkali, 2014(10): 28–30.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Construction Technology of Lightning Protection Devices in Building Electrical Installation Engineering

Xuhong Du*

China Institute for Radiation Protection, Taiyuan 030006, Shanxi Province, China

***Corresponding author:** Xuhong Du, dxh995@163.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: To better improve the electrical safety of buildings and fulfill the role of lightning protection and grounding, this article combines cases and discusses seven aspects of lightning protection devices' construction technology. These aspects include the lightning protection classification of buildings, air terminals, down conductors, grounding devices, division of lightning protection zones, and lightning electromagnetic pulse shielding. Through the introduction of this article, readers can gain a comprehensive understanding of the application and value of lightning protection grounding construction technology in building electrical installation projects.

Keywords: Building electrical; Installation engineering; Lightning protection grounding

Online publication: January 18, 2024

1. Preface

Lightning is a highly destructive natural phenomenon that generates strong currents and high temperatures, causing significant harm to buildings, electrical equipment, communication facilities, and more. In building electrical installation projects, the design and construction of lightning protection grounding systems are crucial. These systems can effectively minimize lightning damage to buildings and electrical equipment [1-3].

As an integral part of construction projects, the design and construction of the lightning protection grounding system in building electrical installations hold immense significance for safeguarding the regular operation of the building and its electrical equipment. The design of the lightning protection grounding system is a critical aspect of building electrical installation engineering. Throughout the design process, factors such as the structural form of the building, the location and model of electrical equipment, soil resistivity, etc., must be considered to determine the optimal lightning protection grounding scheme [4-7].

In general, the lightning protection grounding system comprises components such as air terminals, down conductors, grounding devices, and others. Its design should adhere to relevant standards, such as the "Building Lightning Protection Design Code" (GB50057-2010).

2. Project

2.1. Project overview

The total area of a specific building is 162,847.2 m^2, comprising 61,618.2 m^2 underground and 101,229 m^2 above ground. It consists of two floors below ground and with a floor height of eight meters for two underground floors (4 meters for one floor underground floor). Above ground, there are 18 floors.

2.2. Building lightning protection classification

The lightning protection building designed in this project falls under Class II lightning protection. According to the "Lightning Protection Design Code for Buildings" (GB50057-2010), the classification is determined based on factors such as importance, nature of use, and the potential consequences of lightning accidents. Category II lightning-proof buildings include those with solid or critical electronic information systems, buildings of significant value, and those where the consequences of lightning strikes are severe, necessitating protection.

The building designed for this project is a residential structure with high utility value and personnel density. Moreover, it houses crucial electronic information systems. Following the specified criteria, the building is appropriately classified as a Category II lightning-proof building.

3. Lightning protection and grounding construction of building electrical installation projects

3.1. Lightning protection device

The design content of the lightning protection device in this project encompasses both external and internal components, incorporating measures to prevent the intrusion of lightning surges.

(1) External lightning protection device: To safeguard against direct lightning strikes, the external lightning protection system consists of air terminals, down conductors, and grounding devices [10-15]. Air-termination devices, utilizing φ12 hot-dip galvanized round steel, are strategically placed along the parapet, roof corners, roof ridges, eaves, and areas susceptible to lightning strikes, especially when the entire roof has an air-termination grid smaller than 10 m × 10 m or 12 m × 8 m. When the building exceeds 60 m in height, air-termination strips are initially laid along the roof's perimeter. These strips are positioned on the outer surface of the exterior wall or the vertical surface of the eaves edge. All metal objects protruding from the roof are connected to the roof lightning protection device, while non-metallic objects are equipped with air terminals before connecting to the roof lightning protection device. A circle of down conductors, spaced no more than 18 m apart, surrounds the building perimeter. Diagonal main bars (or steel components) in reinforced concrete columns are full-length welded, connecting the upper end to the lightning strip as well as the lower end to the building foundation beam and steel bar in the foundation floor. Additionally, corresponding to the down conductor used for detection, use a 40 × 4 hot-dip galvanized flat steel below 1.0 m of the outdoor floor to be led to a point greater than 1 m outside the water distribution slope for the connection of dispersion and artificial grounding. Steel bars in the building pile foundation and the foundation beam were used as the grounding device. The two steel bars were welded in the foundation beam with a length of not less than 16 mm as a ring-shaped grounding connection line. The ring-shaped grounding connection line was welded to the two vertical main steel bars in the pile foundation it passes through.

(2) Internal lightning protection device: On the ground floor of the building, the following objects should

be connected to the lightning protection equipotential nearest to the lightning protection device: the building's metal body, metal devices, electrical and electronic systems, and entrances and exits from the building metal pipelines. The external lightning protection device and the metal body of the building, metal devices, and the electrical and electronic systems in the building should meet the separation distance requirements. Among them, in buildings with metal frames, or where steel bars are connected and electrically connected in buildings with reinforced concrete frames, there is no requirement for the separation distance between metal objects or wires and down conductors.

(3) Measures to prevent side lightning strikes: Objects protruding horizontally from the exterior wall, when the ball with a radius of 45 m drops vertically from the roof perimeter lightning strip to the ground and touches the object that protrudes from the exterior wall, corresponding lightning protection measures should be taken. For buildings higher than 60 m, the upper part that accounts for 20% of the height and exceeds 60 m shall be protected against side strikes. The prevention of side strikes shall comply with the following regulations: (i) For the upper part of the building that accounts for 2% of the height and exceeds 60 m, sharp objects, corners, edges, equipment and significantly protruding objects shall be treated according to the protective measures on the roof; (ii) For 20% of the area and exceeding 60 m, the arrangement of air-terminations shall comply with the requirements for this type of lightning protection buildings, and the air-terminations shall be arranged mainly on corners, edges, and significantly protruding objects; (iii) When the minimum dimensions of external metal objects complied with the provisions of the standards, and when external down conductors were arranged at the vertical edges of the building, they can be used as air terminals; (iv) When reinforced concrete inner steel bars and metal frames of buildings met the specifications and used as a down conductor or connected to a down conductor, they can served as an air-termination device; (v) To prevent the intrusion of lightning waves, surge protection devices should be installed at the incoming lines of the power supply system and distribution system; (vi) The distribution lines leading from the distribution box should pass through steel pipes. One end of the steel pipe should be connected to the distribution box and PE line, while the other end should be connected to the electrical equipment's shell and protective cover, with a roof lightning protection device nearby. A jumper is necessary if the steel pipe is disconnected in the middle due to equipment connections.

3.2. Grounding system

The low-voltage power distribution grounding type of this project adopts the TN-S system. The lightning protection grounding and the grounding for strong and weak current systems share a unified grounding plate. The grounding resistance should not exceed 1 Ω. If actual measurements fall short of this requirement, an outdoor artificial grounding electrode should be added, forming a grounding network by welding externally thrown steel bars.

(1) General equipotential bonding measures: Conduct equipotential bonding of protective trunk lines, equipment incoming main pipes, and building metal components inside the building. Establish a total equipotential zone (40 × 4 hot-dip galvanized flat steel) along the interior of the building's exterior wall on the underground floor, connecting all metal pipes entering and exiting the building to it. The general equipotential zone should be securely connected to the grounding device using the main bars in the structural wall and column. Use 40 × 4 thermal galvanized flat steel as the grounding down conductor in distribution rooms, equipment rooms, etc. For each weak motor room, employ two insulated BVR wires for grounding, with the lower end connected to the basic grounding body and the

upper end connected to the grounding terminal box in the machine room. Connect the entire length of metal cable trays and metal brackets to the grounding trunk line at a minimum of two places, with connection points every 20 or 30 m for lengths exceeding 30 m. Ensure reliable connections of the starting and ending ends of the cable tray to the grounding grid. Locally, connect the bottom and top ends of vertically laid metal pipes and metal objects to the grounding device.

(2) Local equipotential bonding (LEB) measures: Install a LEB 0.3 meters above the ground in an appropriate part of the bathroom in the bathing facility. Use two 25 × 4 hot-dip galvanized flat steel pieces to connect the LEB sub-board to the room floor and the mixed side. Connect all metal pipes and metal components in the room to the structural steel bars in the wall using BVR-1 × PC1C pipes, which are then connected to the PE wires of the sockets. Equip general socket circuits with leakage protection switches with a leakage action current of 30 mA. Ensure all lighting fixtures are PE-lined. Establish an equipotential bonding box in the power distribution room. Connect metal devices not charged under normal circumstances to the equipotential junction box, equipment, pipes, components, and other primary metal objects in the building. These connections should reach the nearest lightning protection grounding device or the protective grounding device of electrical equipment. The main equipotential plate is brass, and equipotential connections use various equipotential clips, with welding on metal pipes strictly prohibited.

3.3. Surge protection

Install a Class I test surge protection device (SPD) at the low-voltage power incoming cabinet in the power distribution room. Place another Class I SPD at the LPZ0/LPZ1 boundary, such as the incoming line of the low-voltage distribution cabinet. The SPD's voltage protection level (U_p) should be less than or equal to 2.5 kV. The SPD's impulse current (I_{imp}) of each protection mode should be equal to or greater than 12.5 kA (10/350 μs).

At the LPZ1/LPZ2 boundary, if a Class II test SPD is installed in the distribution box or floor distribution box, the SPD's U_p should be less than or equal to 2.5 kV, the SPD's nominal discharge current (I_n) of each protection mode should be equal to or greater than 20 kA (8/20 μs).

For the third level (LPZ2 and subsequent lightning protection zone boundaries, such as equipment room distribution boxes or information equipment power ports requiring specific protection), install a Class II test or Class III test SPD. Ensure that its I_n is not less than 20 kA (8/20 μs).

4. Summary

Lightning protection grounding technology plays a crucial role in building electrical installation projects, effectively shielding buildings and electrical equipment from potential lightning damage. In environments with frequent lightning activity, structures and equipment are susceptible to strikes, leading to equipment damage and, in extreme cases, personal accidents. Hence, the adoption of effective lightning protection grounding technology becomes imperative to redirect lightning safely into the ground, preventing harm to buildings and electrical equipment.

The implementation of lightning protection and grounding construction technology encompasses various stages, including design, construction, and acceptance. In the design phase, the scheme should consider factors such as the building's structural form, the location and model of electrical equipment, and soil resistivity. During construction, adherence to construction drawings and relevant specifications is essential to ensure that the construction quality meets specified requirements. The acceptance phase involves testing and inspecting the lightning protection grounding system to guarantee its performance aligns with the design requirements and

ensures safe usage.

Future trends in lightning protection grounding technology primarily focus on enhancing the reliability and stability of grounding systems, thereby reducing the likelihood of accidents. Additionally, there is a drive towards developing more efficient and cost-effective lightning protection grounding materials and equipment. The integration of intelligence into lightning protection grounding systems, along with remote monitoring and management capabilities, is gaining importance. Exploring environmentally friendly and sustainable lightning protection and grounding technologies represents another key direction for future development.

Disclosure statement

The author declares no conflict of interest.

References

[1] Wu C, 2023, Lightning Protection and Grounding Construction Technology for Building Electrical Installation Projects. Jiangsu Building Materials, 2023(5): 78–79.

[2] Chen S, Dai Y, Yang X, 2023, Application of Lightning Protection Grounding Technology in Building Electrical Design. Electronic Technology, 52(7): 292–293.

[3] Yang X, You Z, Dai Y, 2023, Application of Lightning Protection Technology in Building Electrical. Electronic Technology, 52(7): 296–297.

[4] Hu J, 2023, Research on Electrical Design of Prefabricated Residential Buildings – Taking Building A as An Example. Real Estate World, 2023(11): 46–48.

[5] Chen W, 2023, Analysis of Design and Construction Technology in Building Electrical Engineering. Integrated Circuit Applications, 40(6): 230–231.

[6] Xu J, 2023, Research on Construction Quality Control of Lighting System in Building Electrical Engineering. Light Source and Lighting, 2023(4): 48–50.

[7] Kong X, Xu F, 2023, Analysis of Installation and Construction Technology of Building Electrical Equipment. China Equipment Engineering, 2023(8): 232–234.

[8] Wu C, 2023, Lightning Protection and Grounding Construction Technology for Building Electrical Installation Projects. Stone Materials, 2023(4): 90–92.

[9] Chen X, 2023, Design Analysis of Power Supply and Distribution System Based on High-Rise Building Electrical Engineering. Jiangxi Building Materials, 2023(3): 130–131.

[10] Li J, 2023, Installation and Construction Technology of Building Electrical Power Supply and Distribution. Jiangsu Building Materials, 2023(1): 88–89.

[11] Shu J, 2023, Lightning Protection Grounding Technology and Its Application in Building Electrical Engineering. China High-Tech, 2023(4): 107–109.

[12] Cui Q, 2023, Analysis on the Lightning Protection Grounding Technology Application in Building Electrical Installation. Sichuan Building Materials, 49(2): 223–224.

[13] Wang L, Wang L, 2022, Lightning Protection and Grounding Design in Building Electrical Engineering. Integrated Circuit Applications, 39(9): 184–185.

[14] China Machinery Industry Federation, 2010, Design Code for Protection of Structures Against Lightning, GB 50057—2010. Ministry of Housing and Urban-Rural Development of the People's Republic of China and General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China

[15] Cao H, Huang X, Mei Y, et al., 2015, Technical Code for Inspection of Lightning Protection System in Building, GB/T 21431—2015. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China and China National Standardization Administration Committee.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

The Impact of Temperature on the Electromagnet and Structural Optimization of a Solenoid Valve

Zhixiong Tang, Zhijun Feng*

School of Mechanical and Automotive Engineering, Guangxi University of Science and Technology, Liuzhou 450200, Guangxi Zhuang Autonomous Region, China

***Corresponding author:** Zhijun Feng, fzjwywmz@163.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The mathematical model of the solenoid valve under varying temperatures is constructed to investigate its performance and enhance heat dissipation balance. The relationship between temperature and electromagnetic force is determined. Electrothermal coupling simulation using COMSOL is conducted, optimizing the outer diameter and length structure parameters of the coil. It is established that the heat dissipation of the coil is influenced by its outer diameter. Subsequently, based on optimized coil structure parameters, an orthogonal experimental design method combined with Ansys Maxwell is employed for simulation solution analysis to study the impact of structural parameters such as length, position, front and rear angles of the magnetic barrier ring in the iron core, armature length, and through-hole size on electromagnetic force. Optimal structural parameters are identified. Results indicate a decrease in steady-state electromagnet temperature by 3–4℃, an increase in the initial electromagnetic force by 32.63%, and a rise in the maximum electromagnetic force by 27.10%.

Keywords: Solenoid valve; Electromagnet; Temperature; Electromagnetic force

Online publication: January 18, 2024

1. Introduction

The solenoid valve, a crucial control element in hydraulic and pneumatic systems, has widespread applications in China's aerospace and navigation sectors. Despite the increasing maturity of solenoid valve design and application in recent years, practical issues such as reduced valve opening capacity at high temperatures or even design defects have been observed [1].

The performance of the electromagnet, serving as the power source for the solenoid valve, plays a decisive role in determining its dynamic characteristics. However, it is worth noting that the coil itself generates heat during operation. As the temperature of the solenoid valve rises, so does its coil resistance, leading to a decrease in current flow. Moreover, thermal stress and expansion within the coil can affect the structure of the solenoid valve. Additionally, environmental factors may lead to wear or even melting of insulating paint on winding coils, ultimately resulting in failure [2-4].

Researchers such as Liu and colleagues as well as Sun and colleagues [5,6], among others, have constructed

physical field coupling models for solenoids and conducted finite element simulation analyses to investigate thermal characteristics under various working conditions, thereby identifying several factors influencing electromagnetic valve temperature distribution. Ren *et al*. studied temperature field distribution through multi-physical field simulation analysis combined with system dynamics hybrid modeling techniques. Their findings revealed significant impacts on dynamic characteristics due to changes caused by temperature fields affecting electromagnet structural parameters [7].

This study focuses on investigating temperature distribution and rise characteristics specific to a company's solenoid valves while proposing reasonable optimization strategies for structural parameters related to both coil and iron core.

2. Mathematical model of solenoid valve

The structure of a single-head electromagnetic directional valve primarily consists of a valve body, valve core, and electromagnet. Its operational principle involves energizing the coil, which generates electromagnetic force to propel the armature of the iron core. Overcoming spring force, frictional resistance, and hydraulic power, the spool moves to open the valve port. Consequently, oil enters through the inlet of the valve port into the flow channel within the valve body and exits from its outlet. Upon de-energizing the coil, electromagnetic force dissipates, and under spring action, the spool resets.

2.1. Equation of mechanical motion

$$\frac{md^2x}{dt^2} = F_m + F_k + F_p + F_f + F_v \qquad (1)$$

Where: m is the quality of the spool; x is spool displacement; t is time; F_m is the electromagnetic force; F_k is the spring force; F_p is hydraulic power; F_f is the friction force on the valve core; F_v is the damping force on the spool.

Based on Maxwell's equations of electromagnetic field, the empirical formula of suction F_m when the leakage flux is ignored is as follows:

$$F_m = \frac{\phi^2}{2\mu_\delta S_\delta} = \frac{B^2 S_\delta}{2\mu_\delta} = \frac{(H\mu_\delta)^2 S_\delta}{2\mu_\delta} = \frac{(IN)^2 S_\delta \mu_\delta}{2\delta^2} = \frac{(\phi R_m)^2 S_\delta \mu_\delta}{2\delta^2} \qquad (2)$$

Where: ϕ is the coil magnetic flux; μ_δ is the permeability of the working air gap; B is the magnetic induction intensity at the working air gap; S_δ is the cross-sectional area of the working air gap; H is the magnetic field strength at the working air gap; N is the number of coil turns; I is the coil working current; δ is the length of working air gap; R_m is magnetoresistance.

2.2. Temperature

From Kirchhoff's voltage law, the current can be derived from the following equation:

$$U = i^2 R + \frac{d(Li)}{dt} \qquad (3)$$

Where: U is the input voltage; i is the input current; R is coil resistance; L is the inductance of the coil.

2.3. Electric field

When the solenoid valve operates with a constant current, the coil generates Joule heat, the formula is as follows:

$$Q_V = \frac{i^2 R_0}{V} \qquad (4)$$

Where: Q_V is the volume heat of coil resistance; R_0 is coil resistance at T_0 temperature; i is the current; V is the coil volume.

Coil resistance is related to conductivity and temperature, and the relationship is as follows:

$$R^{'} = R_0[1 + \frac{(T^{'} - T_0)\gamma_C}{1 + T_0\gamma_c}] \qquad (5)$$

Where: R' is the coil resistance at T' temperature; γ_c is the temperature coefficient of resistivity.

In the case of constant operating voltage, the relationship between the electromagnetic force and temperature is as follows:

$$F_m = \frac{(IN)^2 \mu_\delta S_\delta}{2\delta} = \frac{(UN)^2 \mu_\delta S_\delta}{2\delta (R^{'})^2} = \frac{U^2 N^2 \mu_\delta S_\delta (1 + T_0\gamma_C)^2}{2\delta {R_0}^2 [1 + T_0\gamma_C + (T^{'} - T_0)\gamma_C]^2} \qquad (6)$$

The electromagnetic force, as indicated by **Formula (6)**, is influenced by various factors such as temperature, voltage, resistance, and working air gap. However, the neglect of magnetic leakage in **Formula (6)** leads to inaccurate calculation of the electromagnetic force. Therefore, it becomes imperative to employ the finite element method for comprehending the variations and distribution of electromagnet temperature and assessing the impact of electromagnet structural parameters on the electromagnetic force. Subsequently, optimization can be pursued to enhance solenoid valve performance.

3. Electromagnet temperature simulation

The focus of this paper is primarily on the heating situation of the electromagnet, thus temporarily excluding consideration of the influence of the valve body on heat dissipation. Additionally, this study is based on a specific Vickers solenoid valve as the research subject. A simplified two-dimensional model (as depicted in **Figure 1**) was established in COMAOL to simulate electric-magnetic-temperature field coupling, enabling an understanding of temperature variation and distribution within the electromagnet for conducting reasonable optimization.

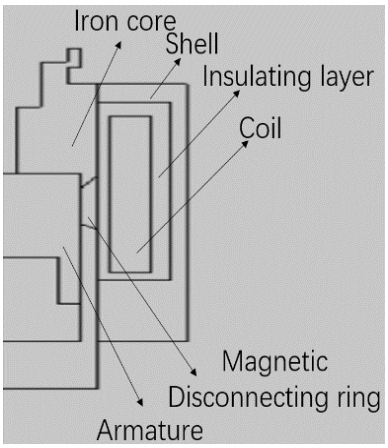


Figure 1. Electromagnet simplified model

According to **Figure 2**, it is evident that the coil winding and both ends of the iron core can reach temperatures as high as approximately 130°C and 100°C, respectively. From **Figure 3**, it can be observed that there is an approximate deviation of about 2°C between simulation and experimental results, while the temperature change curve remains essentially consistent. These findings validate the severity of coil heating and emphasize the necessity to avoid using PWM for voltage adjustment in solenoid valves, ensuring stable current output. Failure of solenoid valves may occur due to power domain reduction under extreme circumstances. Therefore, a comprehensive investigation into factors influencing solenoid temperature becomes imperative.

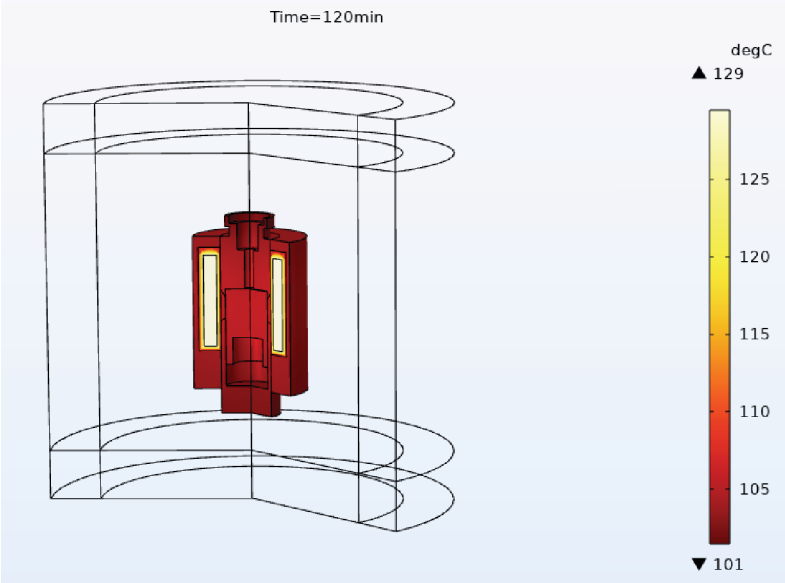


Figure 2. Steady-state temperature distribution of electromagnet

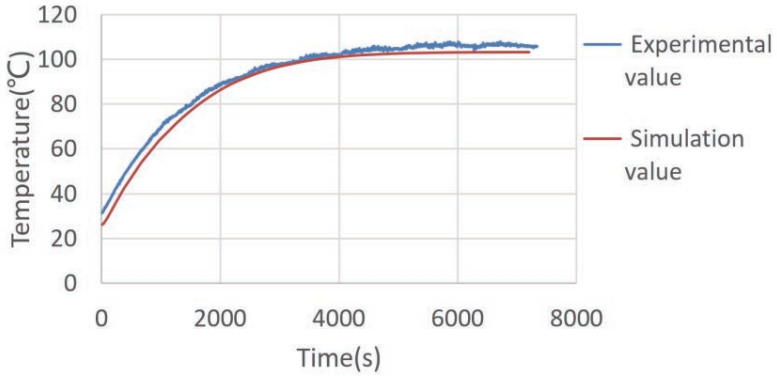


Figure 3. Comparison of experimental and simulated surface temperature of shell side

Numerous factors impact the temperature of the electromagnet, including input voltage, coil resistance, and ambient temperature. This paper focuses on studying and analyzing the structural parameters of the coil as influential factors on temperature while temporarily keeping the size of the iron core unchanged. Consequently, particular attention is given to examining how variations in the outer radius and height of the coil affect its temperature.

The analysis of **Figure 4** indicates that the impact of shell height on electromagnet heat dissipation is not significant. However, an appropriate increase in shell radius can effectively reduce the temperature rise rate and steady-state temperature. Considering the close proximity (2 mm) between the solenoid valve's oil port surface

and coil, a 1.5 mm increment in shell radius is necessary to prevent interference with the installation valve block during operation.

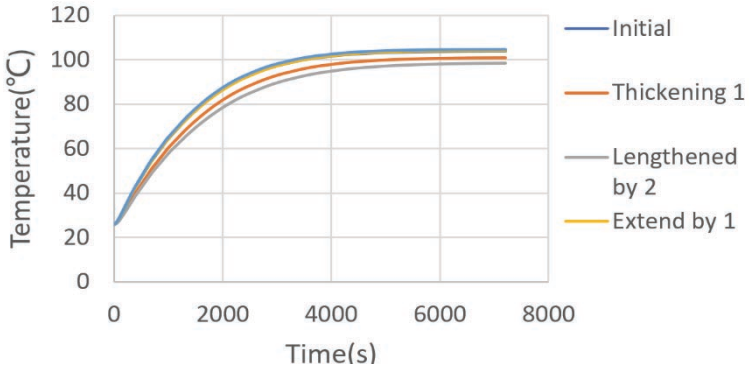


Figure 4. Changes in shell radius and height

4. Electromagnetic force simulation optimization

According to the simulation results and **Formula (6)** mentioned above, increasing the shell radius can effectively reduce the steady-state temperature of the electromagnet while simultaneously enhancing the electromagnetic force when heat dissipation is balanced. To further augment the initial thrust in heat dissipation balance (which corresponds to the furthest effective stroke of the armature), this paper proposes modifying the structural parameters of the iron core. Additionally, based on Ansys Maxwell simulation, **Figure 1** illustrates a simplified model of the electromagnet. The current applied is 0.75A (representing its steady-state value when a solenoid valve is supplied with a voltage of 24 V for simulating operation until heat dissipation reaches equilibrium).

The experimental and simulated magnetic curves in **Figure 5** exhibit a similar trend, albeit with some discrepancies. These errors may arise from the disparity between the simplified model of the electromagnet and its real counterpart, as well as deviations in the relevant parameters of the magnetic material employed in the simulation compared to those of the actual object.

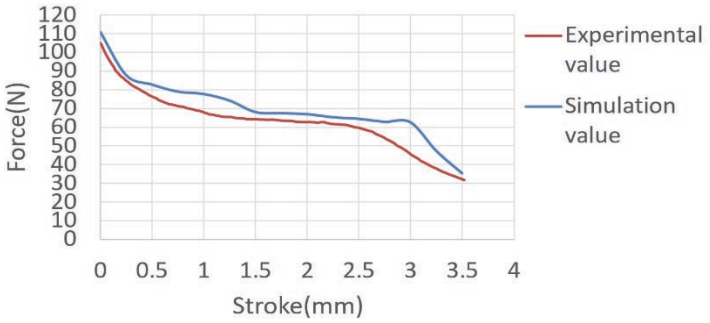


Figure 5. The experimental 0.75 A electromagnetic force curve is compared with the simulated curve

4.1. Electromagnet core optimization

The electromagnetic force is influenced by various factors, including the number of coil turns and the working air gap. Additionally, the magnetic insulation ring, magnetic insulation angle, and structural parameters of the armature also impact the electromagnetic force. Therefore, it is necessary to employ the finite element method for calculation and analysis. The armature serves as a crucial moving component within the iron core, while

the separator ring plays a key role in dividing the magnetic circuit in electromagnets. Consequently, optimizing the structural parameters of both the armature and separator ring primarily constitutes the focus of this paper. The structure diagram of the armature and static core is presented in **Figure 6**, with a selection of 13 factors. These factors include the variation in upper-end length (h1) denoted as 'a,' the variation in lower-end length (h2) denoted as 'b,' the change in bottom hole depth denoted as 'c,' the alteration in diameter of d2 represented by 'd,' the modification in diameter of d1 indicated by 'e,' the adjustment in the number of through-holes referred to as 'f,' and L1 (the distance from the center of d2 to the center of d1) changes designated as 'g.' Additionally, variations such as magnetic separation angles α ('h') and β ('i'), position changes for magnetic barrier ring ('j'), thickness alterations for magnetic barrier ring ('k'), parallel movement changes when face 2 remains stationary ('l'), and parallel movement changes when face 1 remains stationary ('m') were included.

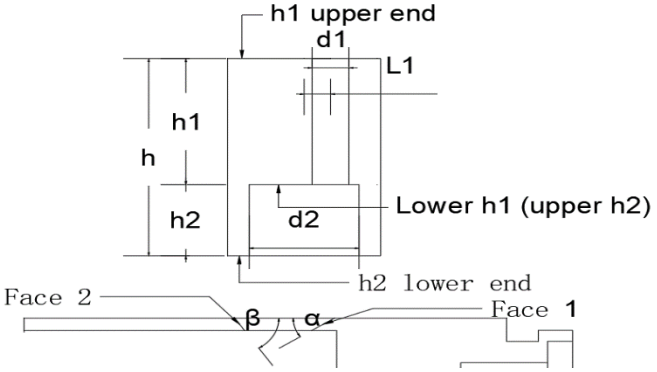


Figure 6. Armature and static core structure

In summary, a total of 13 factors were selected as variables for the optimization of electromagnetic force. Through single-factor analysis, four key factors that significantly affect the magnetic force were identified, namely the position of the magnetic barrier ring, the angle α of the magnetic barrier, and the diameters d1 and d2 of the armature through the armature hole and bottom hole respectively. This study aims to identify a more suitable combination of structural parameters using orthogonal experimental design. Additionally, to enhance oil discharge from the electromagnet armature cavity, only diameter d1 remains unchanged while three levels are chosen for each of the other three factors. Consequently, **Table 1** presents nine combinations arranged accordingly which are then used to calculate the model's electromagnetic force through simulation. Different simulation results have been obtained based on different combinations as shown in **Table 2**.

Table 1. Orthogonal experiment table L_9 (3^3)

Number	e (mm)	h (mm)	j
1	12	Left shift 0.3	36°
2	12	Left shift 0.4	34°
3	12	Left shift 0.5	35°
4	11	Left shift 0.3	34°
5	11	Left shift 0.4	35°
6	11	Left shift 0.5	36°
7	10	Left shift 0.3	35°
8	10	Left shift 0.4	36°
9	10	Left shift 0.5	34°

Table 2. The initial magnetic force value under different combinations

Number	1	2	3
Initial magnetic force (N)	37.473	54.453	54.774
Number	4	5	6
Initial magnetic force (N)	46.249	48.910	49.398
Number	7	8	9
Initial magnetic force (N)	42.507	43.868	60.707

The simulated initial thrust of the original structure is 35.205 N, as depicted in **Figure 5**, indicating that all combinations listed in **Table 2** meet the requirements. However, altering the structural parameters not only affects the initial thrust but also influences the suction force at other positions accordingly. Hence, it is essential to consider whether the electromagnetic force at these points satisfies the requirements for achieving the desired results, as illustrated in **Figure 7**. From **Figure 7**, it can be observed that some position points exhibit a decrease of approximately 2 N in optimized magnetic force compared to the original magnetic force, despite the impact being negligible. Among them, number 1 experiences a minimal increase in initial magnetic force while number 8 surpasses number 7 in terms of initial magnetic force magnitude. Consequently, selecting number 8 leads to a boost of approximately 24.607% in its initial magnetic force without significantly affecting the maximum electromagnetic force.

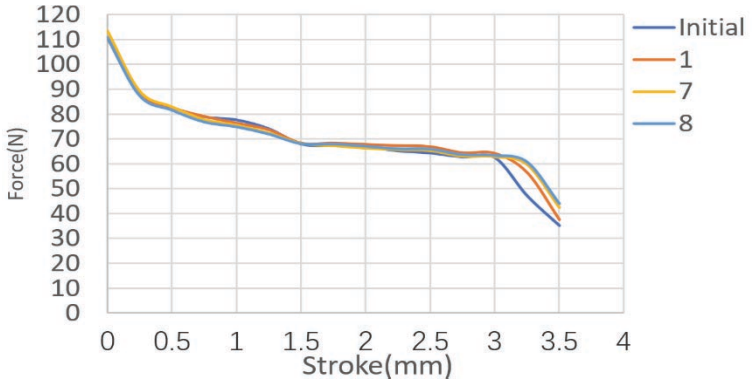


Figure 7. The optimized magnetic curve

5. Conclusion

The temperature field model is established, enabling a comprehensive understanding of the temperature distribution and temperature rise characteristics of the electromagnet. The variation in coil height has a negligible impact on heat dissipation, while an appropriate increase in coil outer diameter can effectively reduce the rate of temperature rise and steady-state temperature. By increasing the coil shell radius by 1.5 mm, its temperature can be reduced by 3–4℃. Through single-factor analysis of the structural parameters of the armature and magnetic barrier ring, four key factors influencing magnetic force are identified: the position of the magnetic barrier ring, magnetic barrier angle α, armature hole diameter d1, and bottom hole diameter d2. Finally, utilizing an orthogonal design table helps identify a set of optimal structural parameters that result in a 24.607% increase in initial magnetic force with minimal change to the maximum electromagnetic force.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Shi Y, Pan Y, 2022, Calculation and Analysis of Influence of Temperature on Solenoid Valve Drive. Mechanical and Electrical Information, 2022(20): 48–51.

[2] Wang Y, Tao G, Yang Y, 2014, Voltage and Temperature Compensation Control Method of Proportional Solenoid Valve. Chinese Hydraulics & Pneumatics, 2014(5): 86–89.

[3] Angadi SV, Jackson RL, Choe S-Y, et al., 2009, Reliability and Life Study of Hydraulic Solenoid Valve. Part 2: Experimental Study. Engineering Failure Analysis, 16(3): 944–963. https://doi.org/10.1016/j.engfailanal.2008.08.012

[4] Li J, Xiao M, Sun Y, et al., 2020, Failure Mechanism Study of Direct Action Solenoid Valve Based on Thermal-Structure Finite Element Model. IEEE Access, 8: 58357–58368. https://doi.org/10.1109/ACCESS.2020.2982941

[5] Liu Y, Mao M, Xu X, et al., 2014, Thermodynamic Analysis of Multi Physical Field Coupling in Hydraulic Solenoid Valves. Journal of Mechanical Engineering, 50(2): 139–145.

[6] Sun B, Liu L, Huang L, et al., 2019, Numerical Simulation and Experimental Research on the Thermal Physical Field of Electromagnetic Valves. Hydraulic and Pneumatic, 2019(5): 92–97.

[7] Ren Y, Xi J, Chen H, et al., 2022, Analysis of Thermal Effects on Proportional Electromagnetic Valves for Vehicles Based on Hybrid Modeling Method. Journal of Beijing Institute of Technology, 42(3): 251–260.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

An Affective EEG Analysis Method Without Feature Engineering

Jian Zhang*, Chunying Fang, Yanghao Wu, Mingjie Chang

Heilongjiang University of Science and Technology, Harbin 150022, Heilongjiang Province, China

Corresponding author: Jian Zhang, zhangmingjian180@qq.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Emotional electroencephalography (EEG) signals are a primary means of recording emotional brain activity. Currently, the most effective methods for analyzing emotional EEG signals involve feature engineering and neural networks. However, neural networks possess a strong ability for automatic feature extraction. Is it possible to discard feature engineering and directly employ neural networks for end-to-end recognition? Based on the characteristics of EEG signals, this paper proposes an end-to-end feature extraction and classification method for a dynamic self-attention network (DySAT). The study reveals significant differences in brain activity patterns associated with different emotions across various experimenters and time periods. The results of this experiment can provide insights into the reasons behind these differences.

Keywords: Dynamic graph classification; Self-attention mechanism; Dynamic self-attention network; SEED dataset

Online publication: January 18, 2024

1. Introduction

The brain, the most complex organ in the human body, has prompted the development of various methods to detect its activity, including electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and more. Among these, EEG has emerged as a primary method for studying brain science due to its high time resolution, cost-effectiveness, and non-invasive characteristics. Emotion, being a conspicuous brain activity, can be objectively reflected through EEG signals. Research on emotion recognition based on EEG signals aims to unveil hidden information within the original EEG data through understanding the relationship between data and tags, thereby discovering the objective law governing brain activity.

Given the high noise and non-stationary characteristics of EEG signals, effective feature extraction methods are crucial for signal classification. Traditional feature extraction methods, such as feature engineering, typically involve preprocessing the data, extracting features (e.g., differential entropy, DE; power spectral density, PSD), and finally employing a simple classifier (e.g., support vector machine, SVM) for classification. Zheng *et al*. utilized DE features and SVM classifiers, achieving an accuracy rate of 83.99% in the full

frequency band using the Shanghai Jiao Tong University (SJTU) Emotion EEG Dataset (SEED) [1]. Additionally, combining DE features with differential asymmetry (DASM), rational asymmetry (RASM), and traditional frequency spectrum (ES) features on symmetrical electrodes proved beneficial [2].

While feature engineering offers good interpretability, directly describing highly abstract features in mathematical language can be challenging. With the recent advancements in deep learning, these features are often extracted automatically with the assistance of neural networks. Zheng *et al*. employed deep belief networks (DBNs), attaining an accuracy of 86.08% across the full frequency band of SEED [1]. Considering the non-Euclidean arrangement of electrode positions in the electrode cap, Song *et al*. applied graph neural networks (GNNs) to emotion recognition, achieving an accuracy of 87.40% by combining graph convolutional neural networks (GCNNs) with DE [3]. Building upon this, the proposed dynamic graph convolutional neural networks (DGCNNs) achieved an accuracy of 90.40%. Furthermore, Liu *et al*. used a three-dimensional convolutional neural network (3D-CNN) based on attention in the time, space, and frequency domains to capture key information between different dimensions of EEG features [4]. Although the accuracy rate reached 96.28%, it is noteworthy that the sample used was 9 seconds, while the aforementioned models typically operated on 1-second intervals.

The hybrid feature extraction method, combining feature engineering and deep learning, has reached a high level. However, given the potent feature extraction capabilities of deep learning, a pertinent question arises: is it possible to set aside feature engineering and directly engage in end-to-end emotion recognition? Kim *et al*. highlighted that EEG signals collected in experimental environments are subject to numerous interference factors, including temporal background, spatial elements, race, language, and culture, resulting in various interference noises in EEG signals [5]. Nevertheless, the end-to-end pattern recognition method is more succinct and offers greater room for development. This holds significant methodological importance for the evolving field of brain interfaces [6].

This paper is based on the dynamic self-attention network (DySAT) proposed by Aravind *et al*. and introduces some enhancements to apply it to emotion recognition in EEG signals [7]. The specific improvements are outlined as follows (https://github.com/zhangmingjian180/DySAT_SEED):

(1) A global pooling layer is added to facilitate a smoother decline in the number of hidden layer units.
(2) The original unidirectional attention in the temporal attention layer is modified to bidirectional attention, enabling the capture of characteristics both before and after.
(3) The sigmoid activation function is replaced with the leaky rectified linear unit (Leaky ReLU) to enhance convergence speed [8].
(4) The residual mechanism is incorporated to stabilize the inter-layer transmission of information [9].

The subsequent sections of this paper unfold as follows: Section 2 briefly introduces the theory related to the model. Section 3 formally presents the DySAT model. Section 4 provides a detailed description of the experiment using DySAT, along with an analysis of the experimental results. Section 5 offers a discussion and a summary of the research in this paper.

2. Related work

The DySAT used in this paper mainly involves two parts: graph attention networks (GATs) and self-attention mechanism [10,11].

2.1. Graph attention network

GAT is essentially a convolution operation, but it is different from the ordinary convolution neural network

(CNN). The latter is based on grid format data, while the former is based on graph, which is a kind of non-Euclidean geometric data. Its formal definition is as follows:

$$G = (V, E) \qquad (1)$$

where V and E represent the set of points and the set of edges. Let the input node eigenvector set be $H = \{h_1, h_2, \ldots, h_n\}$, $h_i \in R^F$, the output node eigenvector set be $H' = \{h'_1, h'_2, \ldots, h'_n\}$, $h'_i \in R^{F'}$, and the parameter set be P, then:

$$H' = \mathrm{GAT}(G, H, P) \qquad (2)$$

2.2. Self-attention mechanism

Self-attention mechanism can be used to extract temporal information from sequences. Previously, this field has been occupied by recurrent neural networks (RNNs). The latest progress of many natural language processing tasks has proved the superiority of the self-attention mechanism in efficiency and performance [11-13]. In principle, GAT is also an example of applying an attention mechanism to graphs.

3. Dynamic graph self-attention network

DySAT is mainly composed of a structural self-attention layer, a sequential self-attention layer, and a global pooling layer. Its overall architecture is shown in **Figure 1**.

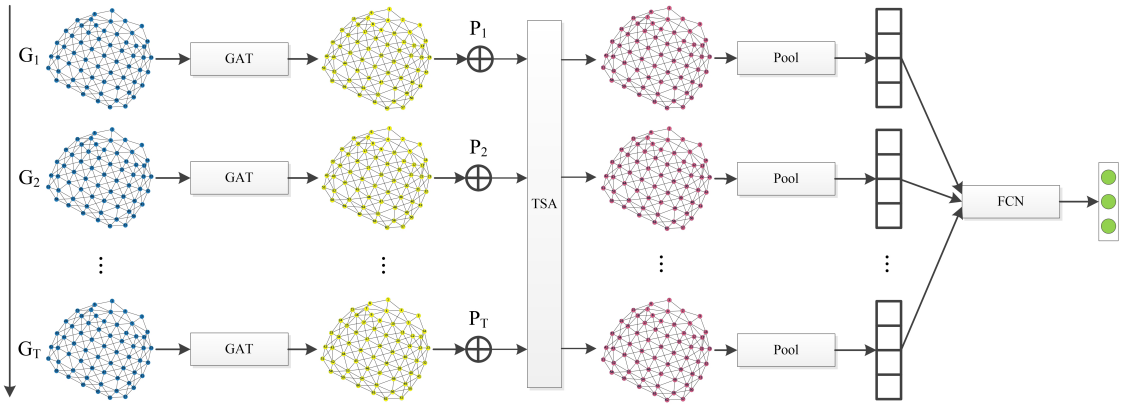


Figure 1. The architecture of DySAT, where GAT represents the graph attention network, TSA represents the tensorized self-attention, and FCN represents the fully connected network

3.1. Model inputs and outputs

The input of the model is a dynamic graph , which is essentially a group of graph sequences, and its formal definition is:

$$D = \{G_1, G_2, \ldots, G_T\} \qquad (3)$$

where T represents the length of the graph sequence. For any $i < j \leqslant T$, meet $G_i.V \subseteq G_i.V$. For a graph G, it can be either a directed graph or an undirected graph. It can also be either a weighted graph or a binary graph.

The output of the model is an n-dimensional vector, and each dimension of the vector represents the corresponding category possibility of the dynamic graph.

3.2. Structural self-attention layer

The structural self-attention layer is used to extract the spatial structure information from dynamic graphs. For the input dynamic graph D, each graph G_i $(1 < i \leqslant T)$ in D undergoes the structure self-attention layer pair,

performing the GAT operation respectively. Consequently, the output of this layer is a dynamic graph D'. The specific calculation processes are as follows:

$$z_v = (\sum_{u \partial N_v} \alpha_{uv} W^s x_u) \qquad (4)$$

$$\alpha_{uv} = \frac{\exp(e_{uv})}{\sum_{w \in N_v} \exp(e_{wv})} \qquad (5)$$

$$e_{uv} = Leaky\ ReLU\ (A_{uv} \cdot a^T [W^s x_u \parallel W^s x_v]) , \forall \langle u, v \rangle \in E \qquad (6)$$

where u,v,w indicate the node number and A_{uv} represents the weight of the edge $\langle u,v \rangle$. For a binary graph, its weight is 1. $N_v = \{v \in V: \langle u,v \rangle \in E\}$. Parameters W^s and α are to be learned for neural networks. *ELU* stands for exponential linear units, *LeakyReLU* stands for leaky rectified linear units. x_u, z_v represents the eigenvector of a node u,v.

To better capture the potential features of different types, the multi-head attention mechanism was used, namely:

$$z_v \leftarrow concat(z_v^1, z_v^2, ..., z_v^{HS}), \forall v \in V \qquad (7)$$

where H_s is the number of heads.

3.3. Sequential self-attention layer

The temporal self-attention layer is used to extract the temporal variation characteristics of nodes. For the input dynamic graph D, temporal self-attention performs the following operations on each node v. The input was set as $X_v = [x^1_v, x^2_v, \ldots x^T_v]^T$, where $x^t_v \in R^{D'}$, whereas the output was set as $Z_v = [z^1_v, z^2_v, \ldots z^T_v]^T$, where $z^t_v \in R^{F'}$. The specific calculation processes are as follows:

Initially, add location embedding:

$$Z_v \leftarrow Z_v + P \qquad (8)$$

where p are the parameters learned through the backpropagation algorithm.

Subsequently, add the self-attention process:

$$Z_v = \beta_v X_v W_v \qquad (9)$$

$$\beta_v^{ij} = \frac{\exp(e_v^{ij})}{\sum_{k=1}^{T} \exp(e_v^{ik})} \qquad (10)$$

$$e_v^{ij} = \frac{(X_v W_q)(X_v W_k)^T}{\sqrt{F}} \qquad (11)$$

where W_v, W_α, W_k are the parameters learned by the backpropagation algorithm.

Here, the multi-head attention mechanism was also used to better capture the potential features of different types:

$$Z_v \leftarrow concat(Z^1_v, Z^2_v, \ldots, Z^{HT}_v), \forall_v \in V \qquad (12)$$

where H_T is the number of heads.

3.4. Global pooling layer

The global pooling layer can simply pool the extracted features, thereby reducing the noise. Simultaneously, it can also smoothen the transformation range of the number of neurons between layers. For the input dynamic

graph D, each graph $G_i(1 \leq i \leq T)$ in the graph D undergoes the pool calculation. The output of this layer $O = [o_1, o_1, ..., o_T\}$. The calculation process of the pool is as follows:

$$o_i = mean\left(h_j \middle| \forall vj \in Gi.V\right) \quad (13)$$

3.5. Loss function

The model is a multi-classification model. Therefore, the classical cross-entropy function was used as the loss function. The calculation expression is as follows:

$$L = -\frac{1}{C}\sum_{i=1}^{C} y_1 \log\left(\frac{e^{\hat{y}_i}}{\sum_j^c e^{\hat{y}_j}}\right) \quad (14)$$

wherein, C represents the number of categories, y represents the sample label, and $\hat{y}$ represents the model output value, both of which are in the form of independent hot coding.

4. Experiment

In this section, the relevant experiments were described using DySAT on SEED [1], followed by an analysis of the experimental results.

4.1. Dataset description

SEED is a public dataset created by the team led by Prof. Bao-Liang Lu and Prof. Wei-Long Zheng in the Brain-Like Computing and Machine Intelligence (BCMI) Laboratory of SJTU. The team utilized 62-lead equipment to sample EEG data from 15 subjects three times each. The sampling frequency was 200 Hz, and a band-pass frequency filter of 0.75 Hz was applied. Subjects were required to watch 15 positive, negative, or neutral movie clips during each sampling session. Before each clip, a 5-second silence period occurred, followed by a ~4-minute clip. Subsequently, subjects conducted a self-assessment for about 45 seconds. After a 15-second rest, subjects proceeded to the next movie clip.

4.2. Introduction to the experiment

Song *et al*. achieved a 90.40% accuracy on the SEED using DGCNN combined with DE features, with a sample length of 1 second [3]. In essence, using 1 second as the sample length theoretically allows for information in the sample to reach at least 90.40% accuracy. Song *et al.*'s study was taken as the theoretical basis for designing the experimental scheme for this study.

Subject-dependent experiments were conducted in this study. The EEG data of each trial for each subject serves as the data unit, resulting in 45 experiments. In each experiment, the EEG data of subjects watching the first nine movie clips form the training set, the next three comprise the validation set, and the final three constitute the testing set. For each emotion, the ratio of training set, validation set, and testing set is 9:3:3.

After dividing the dataset, the data was proceeded for slicing with a length of 1 second. To increase the data volume, the step size of the slice window was set to 0.2 seconds, equivalent to 40 points. The sliced sample data cannot be directly input into the model because the model's input data is a dynamic graph. Therefore, the sample data was converted into a dynamic graph, as described in the subsection. Following this conversion, the data was input into the DySAT model for training and testing, ultimately yielding the experimental results.

4.3. Data preprocessing

In this section, the conversion of the sample data from EEG form to dynamic graph form was introduced, as illustrated in **Figure 2**.

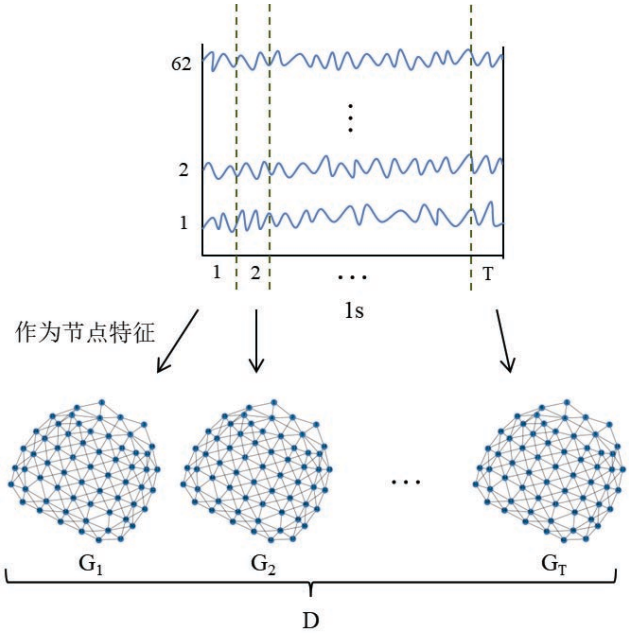


Figure 2. Converting EEG into a dynamic graph

(1) Construction of graphs: This involves creating a binary undirected graph, as depicted in **Figure 3**. In this graph, nodes represent electrodes, and edges indicate adjacency between two electrodes. Consequently, electrode position information is recorded.

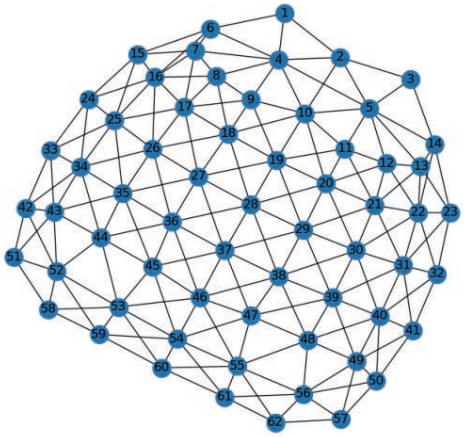


Figure 3. Schematic diagram of the connection between 62 EEG channels, which is used to construct the adjacency matrix of the diagram

(2) Node characteristics: The potential recorded by the electrode was utilized as the node feature. Initially, the sample data was segmented, considering each "small piece" as the state representation of the EEG signal at a specific moment. The reason for this choice, instead of using "every moment" in the real sense, is primarily because EEG signals do not change significantly in a short period of time. Therefore, using a short signal duration better represents the state of the EEG signal. Through repeated experiments, the optimal length for the small signal was determined to be 0.05 seconds, equivalent to

10 data points. This yields a graph sequence length (T) of 20.

(3) Data normalization: Given the substantial difference between the maximum and minimum values of the original potential, direct use as input for the model is impractical, necessitating normalization. According to statistical analysis of EEG signals, the amplitude of EEG signals primarily concentrates between -100 μV to 100 μV. For instance, in the EEG data collected from subject 1 during the first experiment, the highest potential was 14,641.17 μV, while the lowest potential was -12,604.62 μV. The amplitude falls within -100–100 μV, with a ratio of V to V of 0.961. The potential distribution, shown in **Figure 4(a)**, resembles a normal distribution. Similarly, by examining the function image of $f(x) = \tanh(x)$, it was observed that $f(x)$ is approximately linear on (-1,1). To preserve data characteristics to the greatest extent, this paper employs the following formula for data normalization:

$$f(x) = \tanh\left(\frac{x}{100}\right) \quad (15)$$

(a) (b)

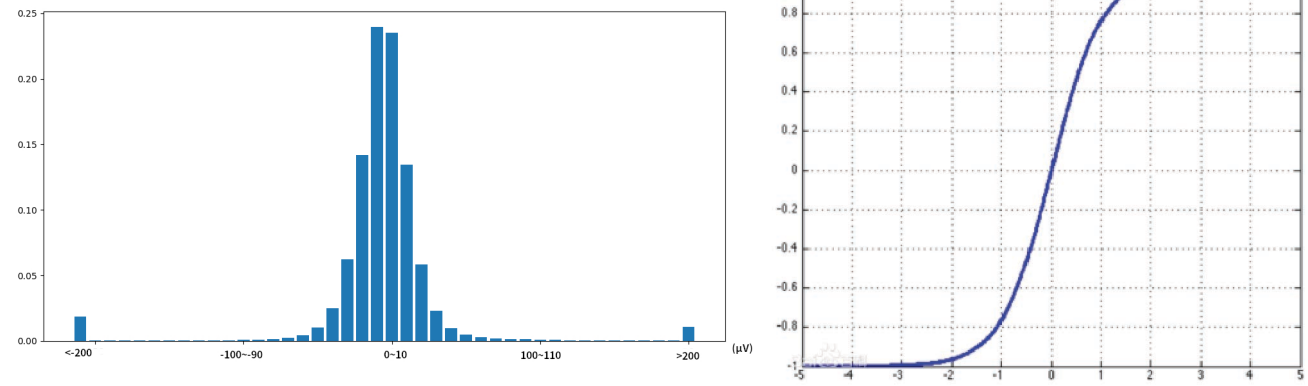


Figure 4. (a) Potential distribution; **(b)** function tanh(x) of subject 1 in the first experiment

4.4. Experimental results and analysis

In deep learning, the configuration of hyperparameters significantly influences model performance. After continuous refinement, a set of approximate optimal hyperparameters consistent with the current experiment was identified. The details are summarized in **Table 1**.

Table 1. Hyperparameters used in the experiment

Name	Value	Explanation
Structural head	2,2	Attention heads in each structural layer
Structural layer	8,8	Units in each structural layer
Temporary head	2,2	Attention heads in each temporary layer
Temporary layer	8,8	Units in each temporary layer
Optimizer	AdamW	Optimizer used in gradient descent
Learning rate	0.001	Initial learning rate for self-attention model
Weight decay	0.000001	This parameter is used in AdamW

Following 45 experiments, the results are presented in **Table 2**. The average accuracy achieved using DySAT as a classifier is 0.6215, markedly higher than random discrimination (with an accuracy of 0.3333 for three classifications). This demonstrates that the end-to-end DySAT, based on a dynamic GNN, can autonomously extract valuable features hidden in EEG data.

Table 2. Experimental results

Subject	Session 1	Session 2	Session 3	AVG	STD
1	0.7224	0.4295	0.4032	0.5184	0.1447
2	0.4041	0.4401	0.5262	0.4568	0.0512
3	0.4960	0.4750	0.3615	0.4442	0.0591
4	0.6796	0.6222	0.4044	0.5687	0.1185
5	0.5202	0.6725	0.6066	0.5998	0.0623
6	0.8655	0.6370	0.7375	0.7467	0.0935
7	0.3432	0.7999	0.3985	0.5138	0.2035
8	0.6544	0.6326	0.8998	0.7289	0.1211
9	0.8034	0.7724	0.5480	0.7080	0.1138
10	0.6240	0.6205	0.5740	0.6062	0.0228
11	0.7224	0.6367	0.8688	0.7426	0.0958
12	0.5398	0.4221	0.7349	0.5656	0.1290
13	0.5380	0.7851	0.8176	0.7136	0.1249
14	0.6559	0.6488	0.5738	0.6262	0.0372
15	0.8274	0.6834	0.8377	0.7828	0.0704
All	All	0.6215	0.1511		

Considering all sessions for all subjects, the standard deviation of the classification results is 0.1511. Across the 45 experiments, the lowest accuracy rate was 0.3432 (Session 1 of subject 7), and the highest accuracy rate was 0.8998 (Session 3 of subject 8). These data reveal significant variations in classification results among different subjects and sessions. Comparing the data in the STD column, it is evident that, except for subject 6, the standard deviation of other subjects is less than the overall average. This indicates that the differences in EEG data between different sessions of the same subject are smaller than those between different sessions of different subjects, emphasizing that subject variation is a crucial factor influencing the characteristics of EEG data.

To align the research in this paper with both domestic and international studies, some data from the experimental results were compared. It is important to note that due to different data preprocessing methods and experimental perspectives (single-mode and multi-mode), the accuracy of experimental results alone cannot conclusively prove the model's performance. Nonetheless, it can offer a reference. **Table 3** indicates that the current effectiveness of emotion recognition algorithms based on feature engineering is not particularly ideal, leaving ample room for improvement. The DySAT model employed in this paper stands out as a superior algorithm among similar approaches.

Table 3. Comparison of similar experiments

Models	Mean	STD	Class	Database
MLP [6]	0.31	0.03	4	AMIGOS
Encoder [6]	0.50	0.01	4	ASCERTAIN
CNN-LSTM [6]	0.37	0.01	4	DECAF
This study	0.62	0.15	3	SEED

5. Summary and discussion

This paper improved DySAT and applied it to emotion recognition in EEG data, demonstrating the model's effectiveness. The experimental results reveal significant variations among different subjects and sessions. It is important to note that these differences are not solely attributable to the model. For instance, in the experiment conducted by Song *et al*. [3], the result was 90.40/8.49, indicating some disparities. Therefore, the primary reasons for the considerable differences in this experiment can be preliminarily attributed to two factors.

Firstly, the inherent nonstationarity of EEG data results in inevitable data differences. However, in theory, appropriately increasing the length of sample data can mitigate this difference. Increasing the sample length theoretically enhances the amount of effective information. Yet, it should be acknowledged that larger samples pose a challenge for the classification method, requiring the handling of a larger volume of data. Balancing these aspects may become a key research direction.

Secondly, differences arise due to limitations within the model itself. As known, the strength of deep learning lies in the "depth" of the model, as deeper models theoretically extract more abstract features. The DySAT proposed in this paper comprises only six layers, which could contribute to the observed difference. It is worth noting that constructing deep GNNs remains an unsolved problem in the field. Anticipated development in GNNs is expected to address this issue.

It is important to recognize that the DySAT proposed in this paper is not the optimal classification model, and there exists a gap between the hybrid feature extraction method, combining feature engineering and deep learning (see Section 2 for details). However, this gap is not fixed, given that deep learning is grounded in the universal approximation theorem [14]. Therefore, the model holds significant potential for improvement. This paper aims to demonstrate the model's effectiveness, leaving the avenue open for substantial performance enhancement in future research.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Zheng W-L, Lu B-L, 2015, Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. IEEE Transactions on Autonomous Mental Development, 7(3): 162–175. https://doi.org/10.1109/TAMD.2015.2431497

[2] Duan R-N, Zhu J-Y, Lu B-L, 2013, 6th International IEEE/EMBS Conference on Neural Engineering (NER), November 6–8, 2013: Differential Entropy Feature for EEG-Based Emotion Classification. IEEE, San Diego, 81–84.

https://doi.org/10.1109/NER.2013.6695876

[3] Song T, Zheng W, Song P, et al., 2020, EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. IEEE Transactions on Affective Computing, 11(3): 532–541. https://doi.org/10.1109/TAFFC.2018.2817622

[4] Liu J, Zhao Y, Wu H, et al., 2021, 13th Asia Pacific Signal and Information Processing Association (APSIPA) Annual Summit and Conference, December 14–17, 2021: Positional-Spectral-Tempotal Attention in 3D Convolutional Neural Networks for EEG Emotion Recognition. APSIPA, Tokyo, 305–312.

[5] Kim J, André E, 2008, Emotion Recognition Based on Physiological Changes in Music Listening. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(12): 2067–2083. https://doi.org/10.1109/TPAMI.2008.26

[6] Dzieżyc M, Gjoreski M, Kazienko P, et al., 2020, Can We Ditch Feature Engineering? End-to-End Deep Learning for Effect Recognition from Physiological Sensor Data. Sensors (Basel), 20(22): 6535. https://doi.org/10.3390/s20226535

[7] Sankar A, Wu Y, Gou L, et al., 2020, The 13th ACM International Conference on Web Search and Data Mining, February 3–7, 2020: DySAT: Deep Neural Representation Learning on Dynamic Graphs via Self-Attention Networks. Association for Computing Machinery, Houston, 519–527. https://doi.org/10.1145/3336191.3371845

[8] Maas AL, Hannun AY, Ng AY, 2013, Proceedings of the 30th International Conference on Machine Learning, June 17–19, 2013: Rectifier Nonlinearities Improve Neural Network Acoustic Models. JMLR, Atlanta, vol. 28, 3.

[9] He K, Zhang X, Ren S, et al., 2016, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27–30, 2016: Deep Residual Learning for Image Recognition. IEEE, Las Vegas, 770–778. https://doi.org/10.1109/CVPR.2016.90

[10] Veličković P, Cucurull G, Casanova A, et al., 2018, 6th International Conference on Learning Representations, April 30–May 3, 2018: Graph Attention Networks. ICLR, Vancouver. https://doi.org/10.48550/arXiv.1710.10903

[11] Vaswani A, Shazeer N, Parmar N, et al., 2017, 31st Conference on Neural Information Processing Systems (NIPS 2017), December 4–9, 2017: Attention Is All You Need. Neural Information Processing Systems Foundation, Long Island. https://doi.org/10.48550/arXiv.1706.03762

[12] Shen T, Jiang J, Zhou T, et al., 2018, DiSAN: Directional Self-Attention Network for RNN/CNN-Free Language Understanding. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1): 5446–5455. https://doi.org/10.1609/aaai.v32i1.11941

[13] Tan Z, Wang M, Xie J, et al., 2018, Deep Semantic Role Labeling with Self-Attention. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1): 4929–4936. https://doi.org/10.1609/aaai.v32i1.11928

[14] Hornik K, Stinchcombe M, White H, 1989, Multilayer Feedforward Networks Are Universal Approximators. Neural Networks, 2(5): 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Research and Application of Artificial Intelligence in Electric Guitar Performance

Tianchang Li*

Orange County School of the Arts, Diamond Bar 91765, California, US

***Corresponding author:** Tianchang Li, frankli9469@gmail.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The article explores the application of artificial intelligence technology in electric guitar playing. It delves into the advantages of artificial intelligence and its seamless integration into electric guitar performance. Additionally, it investigates the application of artificial intelligence technology through an intelligent playing robot. The research aims to offer substantial support for the advancement of artificial intelligence in electric guitar performance.

Keywords: Artificial intelligence technology; Electric guitar playing; Fusion advantage; Guitar-playing robot design

Online publication: January 18, 2024

1. Artificial intelligence and its integration advantages in electric guitar playing

1.1. Overview of artificial intelligence technologies

Artificial intelligence technology, also referred to as AI technology, is a technology designed to learn, study, simulate, and expand human thoughts and behaviors through computer systems and related models, providing ample power for technological innovation. Within the current landscape of AI technology, its primary components encompass image recognition, language recognition, natural language processing, expert systems, and intelligent robots. These technologies have found successful applications in various fields, demonstrating the significant advantages of artificial intelligence.

1.2. Advantages of integrating artificial intelligence technology into electric guitar performance

In contemporary electric guitar performance, the integration of artificial intelligence technology has emerged as a major developmental trend. In this specific integration, the organic fusion of disciplines such as music performance, mechanical design, software engineering, and electronic science enables the electric guitar-playing robot to possess human-like music score recognition abilities and electric guitar-playing skills [1]. Through the cross-fertilization of multiple disciplines and technologies, this integration not only offers robust support for interdisciplinary research in electric guitar music performance but also provides digital and intelligent technical support for modern electric guitar music creation and teaching.

2. The application of artificial intelligence in electric guitar performance research

2.1. Basic overview of electric guitar-playing robots

To rationalize the application of artificial intelligence technology in modern electric guitar playing, this research focuses on studying an electric guitar-playing robot. The primary actuators of the robot responsible for electric guitar playing include the plucking mechanism and the pressing mechanism. Among these, the plucking elements within the plucking mechanism correspond to the number of strings on the electric guitar and operate independently. The plucking strength of each string can be controlled individually. The pressure mechanism comprises a specific number of pressure elements designed to meet the actual pressure requirements of intelligent electric guitar performance. Additionally, the robot is capable of producing the sound of each electric guitar string independently.

2.2. String plucking mechanism design for electric guitar-playing robots

The plucking mechanism of an electric guitar-playing robot consists of two main components.

The first component is the plucked string assembly. Each plectrum assembly comprises a volume control servo, a single plectrum servo, a linkage, a servo arm, and a plectrum swing arm. Specifically designed, the plectrum assembly is configured as a two-by-two counterpart and fixed on the left and right sides of the electric guitar strings. The spacing between each pair of opposing plectrum assemblies should be two strings or more to prevent collisions between the plectrum assembly and the plectrum and pendulum arm. Plucked strings are positioned in the sound hole of the guitar to ensure articulate sound quality. The plectrum swing arm should have functionalities for normal plectrum action, left and right swinging, and servo angle adjustment to achieve optimal sound production, enabling the robot to play a diverse range of electric guitar music [2]. The servo model is set as MG996R, with a supply voltage of 4.8 V and a response speed of 0.17 s per 60° in this mode. For calculating the note duration in music performance, the design follows the formula:

$$t_n = \frac{N}{nT} \quad (1)$$

$$n = 1,2,4,8,16 \quad (2)$$

where N represents the duration of each beat in the piece equal to an n-th note, n is the n-th note, and T is the tempo at which the piece is played in T beats/min.

The second component is the volume adjustment mechanism. In this specific design, pulleys and lines are utilized to adjust the volume of the servo connected to the toggle, which sets the spring on the toggle. Through the toggle, the volume of the servo is adjusted by changing the toggle position when the plucked string pendulum arm is in a particular state, ensuring precise control of the intelligent playing volume.

2.3. Design of string presser mechanism for electric guitar-playing robots

In the string press mechanism of the electric guitar-playing robot, the specific design comprises two main components.

The first component is the string press module, consisting of pressure pads, cylinders, and housing. Three pressure modules are incorporated, each comprising six cylinder actuators with pressure pads corresponding to the six strings on the electric guitar. To accommodate the minimum string spacing of approximately 7 mm, the pressing module utilizes 5.5 mm diameter CJIB4 pin-pad single-acting cylinders. These cylinders have a telescoping length of up to 10 mm and a very rapid response time, enabling effective pressing of the electric guitar strings.

The second component is the moving module, consisting of a linear guide, stepping motor, synchronous

belt, slider, and synchronous wheel. Its primary function is to intelligently control the lateral movement of the string presser module on the fingerboard of the electric guitar. Five linear guide rails are employed, with three power guide rails connected to the string presser module and two support rails providing stability. To adapt to the actual playing mode of the electric guitar, the specific design incorporates a 7 mm wide MGN7 miniature guide rail, corresponding to the slide rail MGN7C type slider, a 6 mm wide GT12 synchronous wheel, and a synchronous belt.

To meet the intelligent playing requirements of electric guitars, the designer must ensure that the travel distance of the string presser module, under the condition of the shortest note-playing time, exceeds the length of the longest demanded value of the guitar's moving fret. The formula for calculating the distance of the presser module's travel for the shortest note-playing time is as follows:

$$L = \frac{n\pi dt}{360T} \qquad (3)$$

where n represents the step angle in the operation of the stepper motor, t represents the pulse period, d represents the diameter of the synchronization wheel, and T represents the time taken for the shortest note to be played.

2.4. Control system design for electric guitar-playing robots

In the design of an electric guitar-playing robot, the control system serves as the primary realization mechanism for artificial intelligence technology. It is through the thoughtful design of this system that artificial intelligence technology can fully leverage its application in modern electric guitar playing.

Within this design of an electric guitar-playing robot, the main components of its control system include a controller, computer, stepper motor driver, servo control module, solenoid valve, and relay. The upper computer functions as the computer, while the lower computer serves as the controller. During specific intelligent control operations, the upper computer primarily analyzes music score data, using it as a basis to plan the movements of the intelligent playing robot. Subsequently, the upper computer transmits the planned data code to the lower computer. Upon receiving the data, the lower computer promptly analyzes it and uses the results as a basis to intelligently drive the robot's string plucking mechanism and string pressing mechanism, thereby executing the intelligent electric guitar performance [3].

Taking the aforementioned design ideas for an intelligent electric guitar-playing robot as a foundation, in this design, the designer initially focused on the hardware of the control system, implementing the following measures:

(1) Controller design: To achieve intelligent control of stepper motors, multiple cylinders, and servos, the designer utilized the STM32RCT6 core board as the controller. The core board boasts a maximum main frequency of 72 MHz, a power supply voltage ranging between 2.0–3.6 V, a program buffer capacity of 256 kB, and 51 inputs and outputs, providing support for 2 interpolated coordinate systems.

(2) Servo control module design: To satisfy the need for simultaneous control of the 12 servos in the intelligent performance robot's structure, the designer employed the Lefthand Sol 16-channel servo control module. This module intelligently drives the servos, facilitating precise manipulation of the 12 servos in the robot.

(3) Stepper motor driver design: To meet the intelligent drive requirements for stepper motors in the robot, the designer selected the TB6600 two-phase stepper motor intelligent driver. This driver operates with a DC input voltage of 9–40 V, an output current range of 0.5–4.0 A, and a maximum

power of 160 W. It supports intelligent drive for both 42-step and 57-step two-phase stepping motors.

(4) Solenoid valve design: To fulfill the demand for intelligent control of single-acting needle cylinders in the pressure string mechanism, the designer opted for the 3V110 two-position, three-normally closed solenoid valve, enabling intelligent control of the cylinder.

Next is the control system software design of the intelligent electric guitar-playing robot. In this specific design, to meet the actual requirements of intelligent control and achieve the desired effect of intelligent electric guitar performance, the designer has specially employed the depth-first search model in artificial intelligence technology. In practical application, this model initially selects a path and searches for the target in one direction, intelligently backtracking when reaching the end of the search. It then selects other paths to continue the search until the target is found. This process enables intelligent control of strumming, pressing the strings, and other actions on the electric guitar, in accordance with specific electric guitar sheet music.

Finally, there is the communication design of the control system for this intelligent electric guitar-playing robot. For the serial communication protocol between the upper and lower computer, the designer primarily employed two types of data frame formats. One is the data frame for the string presser module, used for intelligent planning of the robot's string presser module movement during performance. The other is the data frame for the music score, primarily used for intelligent control of the robot's electric guitar music performance. Regarding the communication protocol in the servo control module, the designer establishes serial communication between the controller and the servo control module. **Table 1** illustrates the data frame format in the servo module of the electric guitar-playing robot.

Table 1. Data frame format in the servo module of this electric guitar-playing robot

Serial number	Enterprise	Specification	Serial number	Enterprise	Specification
1	Header	0*550*55	3	Directives	0*03
2	Data length	L	4	Parameters	Prm1-PrmN

3. Conclusion

In summary, within the realm of electric guitar playing, a playing robot represents a quintessential application of artificial intelligence technology. To meet the practical application demands of this technology, the designer must meticulously design the plucking mechanism, pressure mechanism, and control system of the intelligent playing robot. This ensures the incorporation of intelligent music score recognition functionality, enabling the realization of intelligent electric guitar music performance based on the provided music score. This approach aims to achieve a seamless integration between artificial intelligence technology and electric guitar performance.

Disclosure statement

The author declares no conflict of interest.

References

[1] Zhang Y, 2023, Exploring the Development Road of Fusion of Artificial Intelligence and Music. China Art News.

[2] Zhang K, 2022, Design and Implementation of Artificial Intelligence Curriculum Based on PBL-STEAM Blended Learning Model – Taking Air Guitar Course as an Example. Information and Computer (Theoretical Edition),

2022(5): 251–253.

[3] Sun J, Wen H, 2019, Trial Analysis of the Artificial Intelligence Process of Timbre. Audiovisual, 2019(12): 255–256.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Analysis of Traffic Accidents Based on the Integration Model

Yanshun Ma*, Yi Shi, Yihang Song, Chenxiao Wu, Yuanzhi Liu

The School of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, Heilongjiang Province, China

Corresponding author: Yanshun Ma, dljtxyyhb@163.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: To enhance the safety of road traffic operations, this paper proposed a model based on stacking integrated learning utilizing American road traffic accident statistics. Initially, the process involved data cleaning, transformation, and normalization. Subsequently, various classification models were constructed, including logistic regression, k-nearest neighbors, gradient boosting, decision trees, AdaBoost, and extra trees models. Evaluation metrics such as accuracy, precision, recall, F1 score, and Hamming loss were employed. Upon analysis, the passive-aggressive classifier model exhibited superior comprehensive indices compared to other models. Based on the model's output results, an in-depth examination of the factors influencing traffic accidents was conducted. Additionally, measures and suggestions aimed at reducing the incidence of severe traffic accidents were presented. These findings served as a valuable reference for mitigating the occurrence of traffic accidents.

Keywords: Stacking integrated learning; Data analysis; Traffic safety

Online publication: January 18, 2024

1. Introduction

The elements of road traffic, including people, vehicles, and infrastructure, are experiencing rapid growth, demanding higher standards for road traffic safety. According to the 2022 Statistical Communiqué on National Economic and Social Development by the National Bureau of Statistics, China reported approximately 14,527 deaths due to road traffic accidents in 2022. These accidents have inflicted significant losses to both lives and property, posing serious obstacles to the rapid development of China's transportation industry. Consequently, identifying crucial influencing factors of traffic accidents and predicting their occurrences are of paramount importance and warrant thorough exploration.

In recent years, scholars globally have employed various methods to study influencing factors and predict traffic accidents, yielding notable results. For instance, scholars proposed a logistic regression analysis model, elucidating a significant variable affecting collision severity and analyzing the primary factor contributing to the severity of traffic accidents in Riyadh. Wang *et al.* comprehensively analyzed the influencing factors of urban road traffic accidents from 2014 to 2019, considering human, vehicle,

road, and environmental aspects, and predicted the severity of road traffic accidents using the XGBoost algorithm [1]. Cao *et al.* focused on 6,891 urban road traffic accidents in Heilongjiang Province from 2017 to 2019, introducing the impact of traffic surrounding roads. They employed the Bayesian network to accurately predict severity and influencing factors [2]. Meng *et al.* utilized a logistic regression analysis model to screen factors with significant influence on the consequences of highway traffic accidents, proposing comprehensive measures to mitigate their severity [3]. Chen *et al.* separately applied the C5.0 decision tree algorithm, multivariate logistics regression, and multilayer perceptron (MLP) neural network for prediction. They identified the main influencing factors of traffic accident patterns, providing a basis for decision-making in traffic management departments [4].

2. Data preprocessing

2.1. Selection of characteristic variables

This paper, drawing insights from United States road traffic accident statistics, identifies key variables pertaining to traffic accidents across multiple perspectives. Through a literature research approach, these variables are categorized into three groups: the impact of accidents, the influence of natural factors, and the effects of nearby location and road conditions. Faced with an extensive dataset of quantitative feature variables during the data analysis phase, it becomes imperative to segment this data into non-uniform intervals. Recognizing the limitations of traditional division methods, the study leverages the Pandas and NumPy open-source libraries in Python. The 'pandas.cut( )' method is employed to globally partition the data range, and the 'iloc( )' method visualizes these intervals, generating a set of interval values that constitute the characteristic variable set for the influencing factors of road accidents.

2.2. Data cleansing

To enhance the applicability of the data in model establishment and subsequently improve prediction accuracy, this paper rigorously validates and excludes missing and aberrant data. Additionally, some data undergoes derivative processing. Generally, characteristic variable data can be classified into qualitative and quantitative types. Based on the nature of the data, feature variables related to road accidents are categorized into categorical and numerical variables. To facilitate representation and model application, categorical variables are quantified, and **Table 1** categorizes all feature variables.

Table 1. Summary table of characteristic variable division

Type	Characteristic
Categorical variable	The severity of the accident, weather, day or night, near a speed bump or hump, near the intersection, near the junction, near the location with a no-exit sign, near the railway, near the roundabout, near the station (bus, train, etc.), near stop signs, near a traffic calming device, near traffic signs, whether the position near turning tips, nearby facilities
Numerical variable	Temperature, wind chill, humidity, air pressure, visibility, wind speed, and precipitation

2.3. Data forwarding and normalization

Quantitative data necessitates further processing to eliminate unit influence on model establishment. This paper forwards the quantitative data to standardize dimensions, providing greater convenience in model construction and minimizing the risk of errors. Building on this, this study normalizes the dimensional data, restricting it within the 0–1 range. This ensures more convenient data processing, accelerates program execution, and

enhances the overall efficiency of the entire process.

3. Model building

3.1. Basic principle of a single model

3.1.1. Logistic regression

The logistic regression model serves as a probability model, typically utilizing the likelihood of a variable causing a state as an independent variable that influences factors related to that state. In the context of binary classification logistic regression analysis, where the dependent variable is a binary classification variable (1 indicating occurrence and 0 indicating non-occurrence), if there are N factors associated with it, the probability of the state occurring is calculated as follows:

$$P(Y=1|X)=\frac{1}{1+exp\left[-\left(B_0+\sum_{i=1}^{n}B_ix_i\right)\right]} \quad (1)$$

3.1.2. K-nearest neighbors classifier

The k-nearest neighbors (KNN) classifier relies on data-driven principles. Through statistical screening of the database, it identifies a number of near neighbors with similar features to the current data, conducts statistical analysis on these neighbors, and integrates the results with weights to achieve predictive functions. The KNN method typically begins by structuring the status vector and determining the selection indicator of the nearest neighbor. A distance threshold is set, and Euclidean distance is employed to match the status vector of the current status with the data in the database, thereby obtaining the predictive result of the model. As a classification algorithm, the KNN algorithm boasts advantages such as a clear structure and easy establishment.

3.1.3. Gradient boosting classifier

The gradient boosting classifier is well-suited for regression problems encountered in practical applications. The process of gradient-boosted decision trees (GBDT) is similar to the boosting tree algorithm. However, the distinction lies in the fitting of the loss function. The boosting tree uses the square loss for fitting, while GBDT fits the approximation of the loss in each round by utilizing the negative gradient of the loss function. Subsequently, it fits the regression tree. If the dataset has p data points and q features, and the training dataset is represented as $T=\{(x_1, y_1), (x_2, y_2), \ldots (x_p, y_p)\}$, its loss function can be expressed as $L(y, f(x))$. Initialization involves calculating the negative gradient of the loss function (r_{pi}):

$$\begin{cases} c_{pj}=\sum_{x_i\in R_{pj}} L\left(y_i,\ f_{p-1}(x_i)+c\right) \\ f_p(x)=f_{p-1}(x)+\sum_{j=1}^{J} c_{pj}I\left(x\in R_{pj}\right) \end{cases} \quad (2)$$

Obtaining the regression tree:

$$\widehat{f_{(x)}}=f_P(x)=\sum_{i=1}^{P}\sum_{j=1}^{J} c_{pj}I\left(x\in R_{pj}\right) \quad (3)$$

The gradient boosting tree utilizes the negative gradient of the loss function to fit the approximation of the current round loss, enhancing the model's accuracy and making it more precise than the boosting tree.

3.1.4. Decision tree classifier

The decision tree classifier utilizes extensive datasets to train sample data and articulates classification rules by constructing a tree structure. The decision tree classification model selects the partition attribute by introducing the Gini(D).

$$Gini(D) = 1 - \sum_{k=1}^{K} \left(\frac{|C_k|}{D} \right)^2 \qquad (4)$$

Here, Gini(D) represents the impurity of the data classified according to a specific feature attribute. The smaller the Gini(D), the higher the purity of the data set, indicating less impurity.

3.1.5. AdaBoost classifier

The adaptive boosting (AdaBoost) classifier is an ensemble model that harnesses the idea of the boosting framework. At a deeper level, AdaBoost is an additive model, a method of combining base learners in an additive form. The AdaBoost classification model employs a forward stagewise algorithm and an index loss function algorithm to address the optimization problem associated with the method's addition. The forward stagewise learning algorithm involves continuous iteration through repeated cycles. At each step, it gradually approaches the optimization target formula by considering only one base function, $h_t(x)$, and its coefficient α_t. The index loss function, $E(h(x), y, i) = e^{-yih(xi)}$, contributes to the AdaBoost classification model formula: $H_t(x) = H_{t-1}(x)+\alpha_t h_t(x)$, where h_t represents the t-th base learning device, and α_t is the weight coefficient for each base learning device.

3.1.6. Extra trees classifier

The extra trees (short for extremely randomized trees) classifier is an ensemble learning algorithm that produces classification results by aggregating multiple related outcomes in random forests. Each decision tree is constructed using the original training sample, corresponding to a random sample with each sample having 'k' features. In building each decision tree, the algorithm selects the best features from these characteristic samples and splits the data based on certain indicators. This results in generating random and unrelated decision trees.

3.2. Basic principles of the integration model

The stacking algorithm adopts a layered model integration framework, similar to the bagging algorithm, aiming to enhance accuracy by reducing variance. However, the stacking algorithm achieves improved data fitting by utilizing multiple different models to mitigate variance. Unlike bagging, which trains different data with the same model, stacking employs diverse models to train the same data, resulting in higher accuracy.

The stacking algorithm process begins with the input dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$. This dataset is divided into training and test sets. On the first floor, multiple base learning devices are created using the aforementioned training set. Simultaneously, k-fold cross-validation is performed on the training set, and predictions are made. The features corresponding to each base learning device are collected to form a secondary training set. The test set is then input into the same learning devices to generate the secondary test set. The feature training set is added to the meta-learning device along with the sub-test set. Classification results are predicted, yielding the final prediction model.

The stacking algorithm enhances accuracy by leveraging multiple models for repeated training, resulting in more precise predictions compared to the bagging algorithm. Using the stacking integrated learning method, the seven single models mentioned earlier serve as sub-models, while three linear models function as component models. Three distinct integrated models are derived, and based on the output results, the optimal model is

determined.

4. Model evaluation

4.1. Importance assessment of single model influencing factors

To evaluate the severity recognition rate of traffic accidents using a single model, this article selects various models, including logical regression, KNN classifier, gradient boosting classifier, decision tree classifier, AdaBoost classifier, and extra trees classifier. Five indicators – accuracy, precision, recall, F1-score, and Hamming loss – are employed to determine the optimal solutions among these single models. Higher accuracy rates, precision rates, recall rates, F1-score values, and lower Hamming loss values indicate a better recognition rate for traffic accident severity. To facilitate comparison, line diagrams are drawn for the first four indicators. Each folding line represents a different single model. The horizontal coordinates correspond to the five indicators, while the vertical axis remains within the 0–1 dimensional, as illustrated in **Figure 1**.

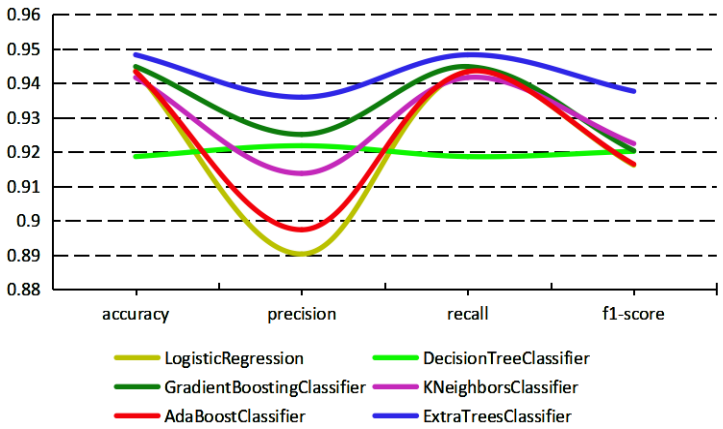


Figure 1. Single model evaluation index line chart

Subsequently, a separate curve chart is drawn for Hamming loss, where the abscissa represents different single models, and the ordinate remains within the 0–1 dimension, as depicted in **Figure 2**.

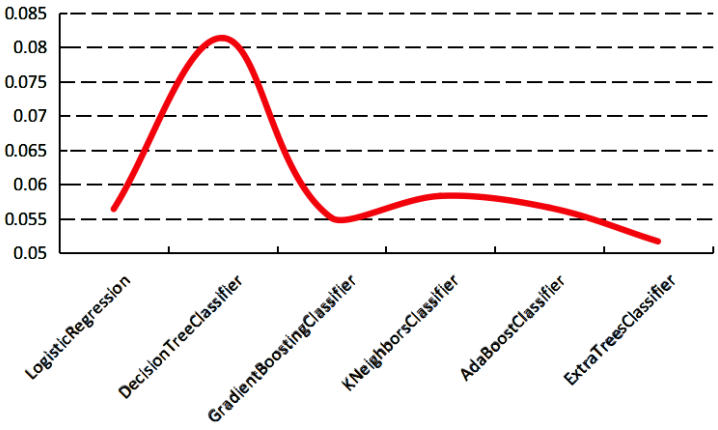


Figure 2. Single model Hamming loss line diagram

Therefore, based on the figures presented, it was concluded that among the single models, the extra trees classifier exhibits the highest recognition rate and the most effective evaluation.

4.2. Assessment of the importance of influencing factors in ensemble models

For evaluating the recognition rate of traffic accident severity using integrated models, this article selects three relatively simple and computationally convenient integrated linear models: passive-aggressive classifier, logistic regression, and ridge classifier. These models serve as the meta-model algorithms in the stacking algorithm, and their process structure is depicted in **Figure 3**.

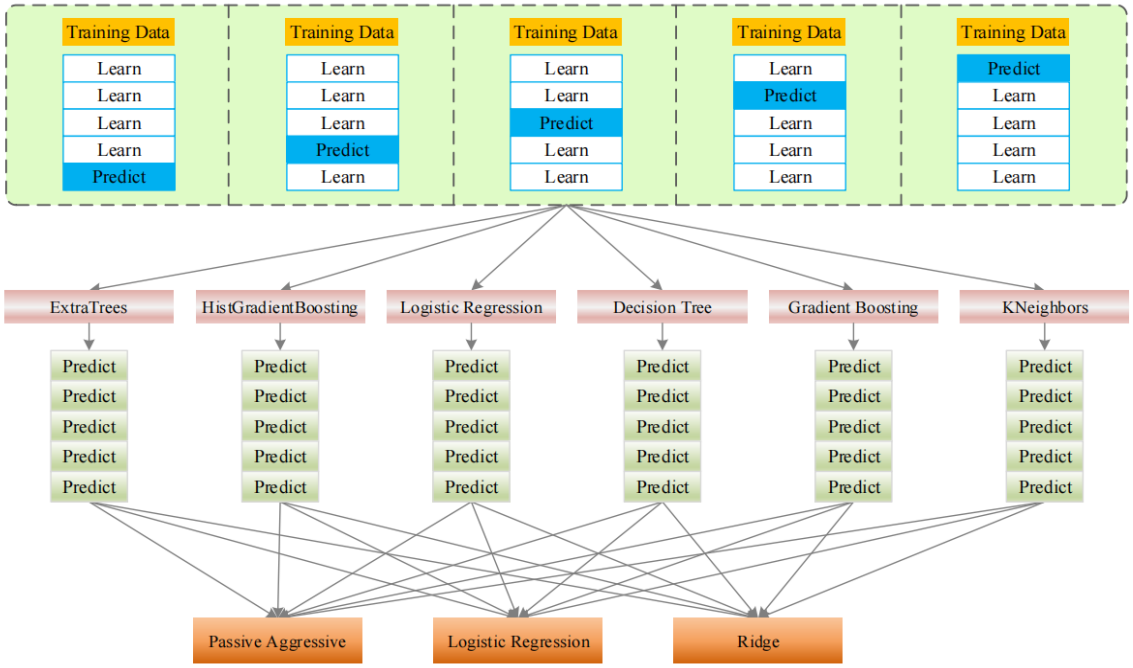


Figure 3. A schematic diagram of the model process structure

Similar to the indicators selected for the impact assessment of a single model, the evaluation of the best scheme among these models is conducted using accuracy, precision, recall, F1 score, and Hamming loss. The Hamming loss is treated separately from other indicators. Using three different lines to represent different integration models, a line chart is drawn for the indicators with values proportional to the recognition rate. The line chart is presented in **Figure 4**.

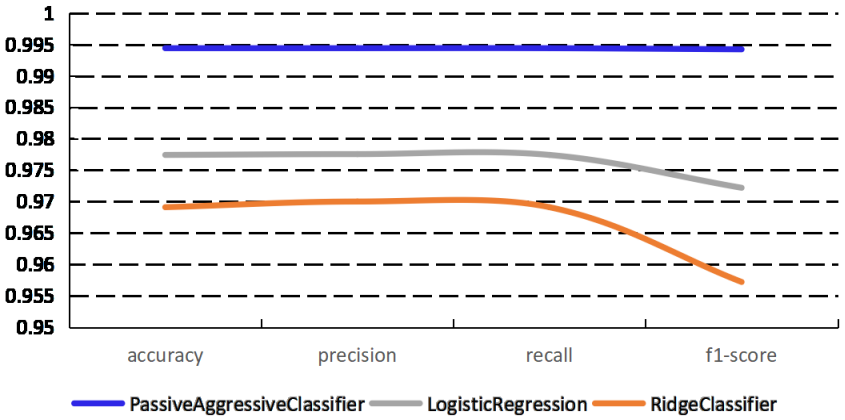


Figure 4. Line diagram of the integrated model evaluation index

Based on **Figure 4**, it is concluded that in the integrated model, the passive-aggressive classifier exhibits the highest recognition rate and the most effective evaluation. Subsequently, the model incorporates the five indicators of weather, time, visibility, nearby convenience facilities, and nearby intersections using permutation feature importance to calculate the importance of the arrangement features that affect traffic accidents for the optimal model.

Table 2. Characteristics and corresponding importance weights

Weight	Feature	Weight	Feature
0.0473 ± 0.0004	Humidity (%)	0.0047 ± 0.0002	Station
0.0453 ± 0.0002	Pressure (in)	0.0045 ± 0.0001	Precipitation (in)
0.0441 ± 0.0002	Temperature (F)	0.0034 ± 0.0001	Stop
0.0429 ± 0.0001	Wind_Chill (F)	0.0012 ± 0.0001	Railway
0.0397 ± 0.0003	Wind_Speed (mph)	0.0008 ± 0.0000	Junction
0.0336 ± 0.0002	Weather_Condition	0.0003 ± 0.0000	No_Exit
0.0225 ± 0.0004	Sunrise_Sunset	0.0003 ± 0.0000	Give_Way
0.0203 ± 0.0003	Traffic_Signal	0.0001 ± 0.0000	Traffic_Calming
0.0144 ± 0.0001	Crossing	0.0000 ± 0.0000	Bump
0.0128 ± 0.0002	Visibility (mi)	0.0000 ± 0.0000	Roundabout

5. Countermeasures analysis

5.1. Existing problems

The road traffic system, comprising people, vehicles, roads, and the environment, forms a complex dynamic system. The causes of traffic accidents primarily encompass subjective factors, such as psychological or physiological aspects of the involved parties, and objective factors, including inherent attributes of the vehicle like weight and configuration, road geometric composition (e.g., number of lanes, driving sight distance), changing areas of road environmental characteristics, and management factors.

Based on the analysis of the significance of arrangement characteristics affected by the above outputs, environmental factors emerge as having a substantial impact on the severity of traffic accidents. These factors include temperature, humidity, weather conditions, pressure, light, and visibility. In recent years, the frequency of traffic accidents caused by meteorological disasters has risen, with adverse weather conditions being a primary factor affecting the safe operation of the highway network. Results indicate that the structural composition, road facilities, and the road's environment significantly influence road traffic safety. Factors such as the presence of intersections, signal lights, lighting settings, and traffic signs near the road all contribute to the occurrence of traffic accidents.

5.2. Countermeasures to promote traffic safety

5.2.1. Environmental aspects

Survey data and statistics reveal that 25% of highway accidents in China are attributed to foggy weather, with nearly one-third of major traffic accidents linked to such conditions. In response, some countries may close highways during heavy fog, and drivers on other highways should promptly turn on their fog lights, maintain a safe following distance, and reduce driving speed. In addition to foggy weather, ice and snow conditions can significantly impact road operations. Drivers should stay informed about road conditions during snowy weather,

avoid forcing passage in heavy snow, and attempt detours if necessary. Transportation departments should swiftly organize resources to clear snow and implement vehicle deceleration facilities to ensure road traffic safety in extreme weather. Precipitation also affects the severity of traffic accidents, so road structures should accommodate maximum drainage capacity to ensure driving safety during rainy and snowy weather.

5.2.2. Road aspects

The characteristics of road sections also influence the severity of traffic accidents. Environmental factors near road sections, such as intersections, stations, and railways, have a notable impact on accident severity. Intersection sections show a higher frequency of serious accidents compared to general sections, and sections near stations and railways also exhibit a more prominent frequency of serious accidents. Installing traffic management facilities and equipment in high-risk road sections is significant for reducing the incidence of serious traffic accidents.

5.2.3. Management aspects

To ensure the safe operation of road traffic, traffic management departments, public security organizations, highway bureaus, and other relevant units employ education, technology, and other means to restrict traffic reasonably, organize and command traffic scientifically, and manage the relationship between people, vehicles, and roads correctly. Improving road traffic management facilities, such as installing traffic signs and lights, aids road users in safely navigating through road sections. With the advancement of science and technology, the utilization of information-based intelligent transportation meets the evolving needs for safe, orderly, smooth, and sustainable transportation.

6. Conclusion

With the acceleration of urbanization, the complexity of traffic composition is on the rise, leading to a significant increase in the probability of traffic accidents compared to previous times. This article employs ensemble learning methods in machine learning to stack multiple models that are relatively effective in predicting traffic accident risk and applies a simple linear method. The result is a fusion model with high-performance prediction for traffic accident risk. It achieves an accuracy of 0.994 for public data, a recall of 0.994, an F1 score of 0.99423, and a Hamming loss of 0.0056. Conducting a permutation feature importance analysis on the model identifies the characteristics that have a significant impact on traffic accident risk. Based on these important features, corresponding suggestions and improvement measures are proposed. The research conclusions can serve as a reference for further studies.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Wang Y, Liu Q, Cheng W, 2022, Prediction of Road Traffic Accident Severity Based on the XGBoost Algorithm. Software Guide, 21(5): 84–88.

[2] Cao Y, Zhang B, Li S, 2022, Analysis of the Factors Influencing the Severity of Urban Road Traffic Accidents in the Ice and Snow Season. Journal of Dalian Jiao Tong University, 43(4): 8–13.

[3] Meng Y, Zhang X, Qing G, et al., 2022, Analysis of Influencing Factors on the Consequences of Highway Traffic Accidents Based on Logistic Regression. Journal of Wuhan University of Technology (Transportation Science and Engineering Edition), 46(1): 12–16.

[4] Chen L, Li C, Zhan L, et al., 2022, Analysis and Prediction of Urban Road Traffic Accidents. Journal of Chang'an University (Natural Science Edition), 42(4): 98–107.

[5] Zhu H, Li H, Chi Y, et al., 2022, A Method for White Layer Prediction of Hard Car Surface Based on Gradient Lifting Decision Tree. Shanghai Municipality, CN115099266A.

[6] Liu G, Gao Y, 2021, Review of Intrusion Detection Methods Based on Machine Learning. Information and Computer (Theoretical Edition), 33(10): 34–37.

[7] Zhang P, Xu S, 2008, Exploring the Method of Road Traffic Safety Management and Management Planning in Heilongjiang Reclamation Area. Heilongjiang Science and Technology Information, 2008(19): 82.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Comprehensive Benefit Evaluation of SZ Distributed Photovoltaic Power Generation Project Based on AHP-Matter-Element Extension Model

Shuli Jing*

Business School, Shandong University of Technology, Zibo 255049, China

***Corresponding author:** Shuli Jing, jingshulirw@126.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: With the introduction of the "dual carbon goals," there has been a robust development of distributed photovoltaic power generation projects in the promotion of their construction. As part of this initiative, a comprehensive and systematic analysis has been conducted to study the overall benefits of photovoltaic power generation projects. The evaluation process encompasses economic, technical, environmental, and social aspects, providing corresponding analysis methods and data references. Furthermore, targeted countermeasures and suggestions are proposed, signifying the research's importance for the construction and development of subsequent distributed photovoltaic power generation projects.

Keywords: Distributed photovoltaic power generation; Comprehensive benefits; Evaluation

Online publication: January 18, 2024

1. Introduction

The comprehensive evaluation of distributed photovoltaic power generation projects enables a systematic assessment of their economic, technical, environmental, and social benefits. By summarizing the experiences and practices derived from completed projects, potential issues and shortcomings in ongoing project constructions can be promptly identified. This article focuses on the SZ distributed photovoltaic power generation project, examining its economic, technical, environmental, and social benefits. A more scientifically structured comprehensive benefit evaluation index system is introduced, utilizing the AHP-matter-element extension model for in-depth analysis. This approach not only provides insights into the project's overall benefits but also serves as a reference for future photovoltaic power generation initiatives. The study concludes by offering targeted countermeasures and suggestions to enhance the comprehensive benefits of distributed photovoltaic power generation projects.

2. Overview of SZ distributed photovoltaic power generation project

The focus of this research is the SZ distributed photovoltaic power generation project. The project utilizes

a building roof area of 410 m^2, and to ensure both roof safety and user capacity, the planned investment and construction scale for the photovoltaic power station is 47.4 kWp. This involves a 50 kW photovoltaic inverter, a grid-connected cabinet, and integration into the user-side power distribution system at a 400 V voltage level.

The estimated total power generation over 25 years is 1.6 million kWh, averaging an annual production of 64,000 kWh. The grid-connected approach emphasizes self-consumption, with the excess electricity (30%) being fed back into the grid. The project follows the EPC general contracting method, excluding photovoltaic components, inverters, photovoltaic grid cabinets, photovoltaic project cables, and photovoltaic brackets. The expected investment return rate stands at 7.69%.

3. Construction of a comprehensive benefit evaluation index system for SZ distributed photovoltaic power generation projects

3.1. Principles of index system construction

Constructing a comprehensive benefit evaluation index system for distributed photovoltaic power generation projects necessitates consideration of the project's economic, technical, environmental, and social benefits. Simultaneously, adherence to principles of scientificity, systematicness, operability, and the integration of quantitative and qualitative evaluation is essential. This ensures the accuracy and objectivity of the evaluation results [1].

3.2. Determination of evaluation indicators

The selection of evaluation objects should be based on the specific conditions of the distributed photovoltaic power generation project and the evaluation objectives. The evaluation object could be the entire project, a specific project component, or a particular indicator. In choosing the evaluation object, its representativeness and importance must be thoroughly considered to ensure the evaluation results reflect the overall comprehensive benefits of the project [2]. By collating and analyzing current research results, considering the actual situation of the SZ distributed photovoltaic power generation project, and consulting expert opinions, economic benefits and technical benefits are determined. Evaluation indicators are then selected from four aspects: economic benefits, technical benefits, environmental benefits, and social benefits, resulting in the construction of a comprehensive benefit evaluation index system for SZ distributed photovoltaic power generation projects, as depicted in **Table 1**.

3.2.1. Economic benefits

In distributed photovoltaic power generation projects, economic benefits take precedence. The project's successful operation must ensure that the generated income is sufficient to repay short-term and long-term debts [3]. Primary indicators are derived from two aspects: profitability and solvency. Profitability indicators encompass three secondary indicators: investment payback period, net capital profit rate, and total investment return rate. Solvency indicators include two secondary indicators: the asset-liability ratio and the liquidity ratio.

3.2.2. Technical benefits

The quality of the selected project site significantly influences the size and efficiency of the solar energy resources that the project can effectively harness. Simultaneously, the hardware's performance and stability directly impact the project's final operational effectiveness [4]. First-level indicators are predominantly determined by project site selection and system performance. Site selection indicators encompass two secondary indicators: site selection conditions as well as solar radiation and climate conditions. System performance indicators include two secondary indicators: photovoltaic array design as well as photovoltaic system efficiency and reliability [5].

Table 1. Comprehensive benefit evaluation index system for SZ distributed photovoltaic power generation projects

Target	Criterion layer	First level indicator	Secondary indicators
SZ distributed photovoltaic power generation project comprehensive benefit Z	Economic benefit A_1	Profitability B_1	Payback period C_1
			Net profit margin on capital C_2
			Total investment return C_3
		Solvency B_2	Asset-liability ratio C_4
			Current ratio C_5
	Technical benefit A_2	Project site selection B_3	Site selection condition C_6
			Solar radiation and climate conditions C_7
		System performance B_4	Photovoltaic array design C_8
			Photovoltaic system efficiency and reliability C_9
	Environmental benefits A_3	Energy saving and emission reduction B_5	Average annual coal emission reduction C_{10}
			Average annual carbon dioxide emission reduction C_{11}
			Average annual sulfur dioxide emission reduction C_{12}
			Average annual carbon and nitrogen emission reduction C_{13}
		Impact on regional environment B_6	Land occupation reduction C_{14}
			Ecological environment improvement C_{15}
	Social benefit A_4	Socioeconomic B_7	Economic development promotion C_{16}
			Increased jobs C_{17}
		Social environment B_8	Energy structure improvement C_{18}
			Public awareness C_{19}

3.2.3. Environmental benefits

Photovoltaic power generation, as a clean energy source, contributes substantially to environmental benefits. Distributed photovoltaic power generation projects play a crucial role in reshaping the current energy structure, reducing reliance on traditional energy sources, and curbing toxic and harmful gas emissions. First-level indicators are primarily determined by energy conservation, emission reduction, and their impact on the regional environment. Energy conservation and emission reduction indicators include average annual coal emission reduction, annual carbon dioxide emission reduction, annual sulfur emission reduction, and annual average carbon and nitrogen reduction, totaling four secondary indicators. The impact on the regional environment includes two secondary indicators: reducing land occupation and improving the ecological environment.

3.2.4. Social benefits

Lastly, social benefits encompass a broad spectrum. Distributed photovoltaic power generation projects not only create employment opportunities, alleviating employment pressure to some extent, but also stimulate the development of related local industries. First-level indicators are mainly determined by two aspects: socioeconomic and the social environment. Socioeconomic indicators include two facets: promoting economic development and increasing job opportunities. Social environment indicators encompass two aspects: improving the energy structure and enhancing social public awareness.

4. Comprehensive benefit evaluation of SZ distributed photovoltaic power generation project

According to the constructed comprehensive benefit evaluation index system, the analytic hierarchy process (AHP) is utilized to calculate the weight of the indicators. Additionally, a matter-element extension model of the comprehensive benefit of the SZ distributed photovoltaic power generation project is developed to evaluate the overall project benefits.

4.1. Determination of indicator weight based on the AHP method

4.1.1. Construct a judgment matrix

Referring to the indicator system outlined in **Table 1**, indicators at the same level are systematically compared in pairs. Expert scoring, combined with the scale values from **Table 2**, determines the degree and importance of their impact on the upper-layer indicators. This information is expressed in matrix form to generate a judgment matrix.

Table 2. The scale and meaning of each element in the judgment matrix

Scale value	Judgment of pairwise comparison results
1	The i element and the j element are equally important.
3	The i element is slightly more important than the j element.
5	The i element is significantly more important than the j element.
7	The i element is strongly more important than the j element.
9	The i element is more important than the j element.
2, 4, 6, 8	Indicates the intermediate degree of the above adjacent judgments
The reciprocal of the above scale value	The importance coefficient of element j relative to element i

4.1.2. Calculation of index weight and consistency test

The evaluation index system and the constructed judgment matrix undergo a consistency test using MATLAB software. Once the test is successfully passed, the results for the index weights are calculated, as presented in **Table 3**.

Table 3. Weight table of comprehensive benefit evaluation index system for SZ distributed photovoltaic power generation projects

Target	Criterion layer	Weight	First level indicator	Weight	Secondary indicators	Weight
SZ distributed photovoltaic power generation project comprehensive benefit Z	Economic benefit A_1	0.5638	Profitability B_1	0.7500	Payback period C_1	0.0516
					Net profit margin on capital C_2	0.1351
					Total investment return C_3	0.2361
			Solvency B_2	0.2500	Asset-liability ratio C_4	0.1175
					Current ratio C_5	0.0235
	Technical benefit A_2	0.2634	Project site selection B_3	0.2000	Site selection condition C_6	0.0395
					Solar radiation and climate conditions C_7	0.0132
			System performance B_4	0.8000	Photovoltaic array design C_8	0.0702
					Photovoltaic system efficiency and reliability C_9	0.1405
	Environmental benefits A_3	0.1178	Energy saving and emission reduction B_5	0.8000	Average annual coal emission reduction C_{10}	0.0371
					Average annual carbon dioxide emission reduction C_{11}	0.0220
					Average annual sulfur dioxide emission reduction C_{12}	0.0220
					Average annual carbon and nitrogen emission reduction C_{13}	0.0131
			Impact on regional environment B_6	0.2000	Land occupation reduction C_{14}	0.0059
					Ecological environment improvement C_{15}	0.0177
	Social benefit A_4	0.0550	Socioeconomic B_7	0.7500	Economic development promotion C_{16}	0.0330
					Increased jobs C_{17}	0.0083
			Social environment B_8	0.2500	Energy structure improvement C_{18}	0.0115
					Public awareness C_{19}	0.023

4.2. Comprehensive benefit evaluation of SZ distributed photovoltaic power generation project based on matter-element extension model

The matter-element analysis method, initially proposed by Professor Cai Wen, provides a solution to the incompatibility problem from both qualitative and quantitative perspectives. Comprehensive benefits analysis of distributed photovoltaic power generation projects encompasses multiple dimensions including economy, technology, environment, and society. The influence of evaluation indicators on the results varies. The matter-element extension model adeptly combines qualitative and quantitative indicators, accurately describing the comprehensive benefits and the mutual influence and interaction among various evaluation indicators in the evaluation index system [6]. The model-building steps are as follows:

4.2.1. Determine the classical domain, nodal domain, and matter elements to be evaluated

$$R_j = (N_j, C_i, V_{ji}) = \begin{bmatrix} N_j & C_1 & V_{j1} \\ & C_2 & V_{j2} \\ & \vdots & \vdots \\ & C_n & V_{jn} \end{bmatrix} = \begin{bmatrix} N_j & C_1 & [a_{j1}, b_{j1}] \\ & C_2 & [a_{j2}, b_{j2}] \\ & \vdots & \vdots \\ & C_n & [a_{jn}, b_{jn}] \end{bmatrix} \quad (1)$$

In **Formula (1)**, N_j represents the j-th evaluation level of the rating object, C_i represents the i-th evaluation index of N_j, V_{ji} represents the value range of N_j with respect to C_i, and $[a_{ji,}\ b_{ji}]$ is the classic domain.

$$R_p = (N_p, C_i, V_{pi}) = \begin{bmatrix} N_p & C_1 & V_{p1} \\ & C_2 & V_{p2} \\ & \vdots & \vdots \\ & C_n & V_{pn} \end{bmatrix} = \begin{bmatrix} N_p & C_1 & [a_{p1}, b_{p1}] \\ & C_2 & [a_{p2}, b_{p2}] \\ & \vdots & \vdots \\ & C_n & [a_{pn}, b_{pn}] \end{bmatrix} \quad (2)$$

In **Formula (2)**, N_p represents all levels of the index to be evaluated, V_{pi} represents the set of value ranges of N_p with respect to all levels of ,C_i and $[a_{pi,}\ b_{pi}]$ is the node domain.

$$R_0 = (N_0, C_i, V_i) = \begin{bmatrix} N_0 & C_1 & V_1 \\ & C_2 & V_2 \\ & \vdots & \vdots \\ & C_n & V_n \end{bmatrix} \quad (3)$$

In **Formula (3)**, R_0 represents the matter element to be evaluated, N_0 represents the evaluation index, and V_i represents the measured data of the characteristic value of N_0 with respect to C_i.

Evaluation indicators are classified into five levels: “poor, relatively poor, average, good, and excellent.” Quantitative indicators’ classical and sectional domains are determined by referring to relevant industry standards, existing research results, and consulting expert opinions. Qualitative indicators are categorized into five levels based on project characteristics and converted into numerical values, with the value range being [0, 100]. The classical and nodal domains are then determined based on the project conditions.

Table 4. Evaluation indicators classic domain and section domain

Index	Poor	Relatively poor	Average	Good	Excellent	Joint domain
C_1	[16, 25]	[13, 16]	[10, 13]	[7, 10]	[0, 7]	[0, 25]
C_2	[0, 1]	[1, 6]	[6, 12]	[12, 18]	[18, 27]	[0, 27]
C_3	[0, 2]	[2, 3.9]	[3.9, 4.6]	[4.6, 5.7]	[5.7, 8]	[0, 8]
C_4	[73.1, 77.7]	[66.1, 73.1]	[61, 66.1]	[55.1, 61]	[0, 55.1]	[0, 77.7]
C_5	[0, 0.5]	[0.5, 1]	[1, 1.5]	[1.5, 2]	[2, 3]	[0, 3]
C_6	[0, 20]	[20, 40]	[40, 60]	[60, 80]	[80, 100]	[0, 100]
C_7	[0, 20]	[20, 40]	[40, 60]	[60, 80]	[80, 100]	[0, 100]
C_8	[0, 20]	[20, 40]	[40, 60]	[60, 80]	[80, 100]	[0, 100]
C_9	[0, 20]	[20, 40]	[40, 60]	[60, 80]	[80, 100]	[0, 100]
C_{10}	[0, 5]	[5, 10]	[10, 15]	[15, 20]	[20, 25]	[0, 25]
C_{11}	[0, 15]	[15, 30]	[30, 45]	[45, 60]	[60, 75]	[0, 75]
C_{12}	[0, 0.05]	[0.05, 0.1]	[0.1, 0.15]	[0.15, 0.2]	[0.2, 0.25]	[0, 0.25]
C_{13}	[0, 0.04]	[0.04, 0.08]	[0.08, 0.12]	[0.12, 0.16]	[0.16, 0.2]	[0, 0.2]
C_{14}	[0, 20]	[20, 40]	[40, 60]	[60, 80]	[80, 100]	[0, 100]
C_{15}	[0, 20]	[20, 40]	[40, 60]	[60, 80]	[80, 100]	[0, 100]
C_{16}	[0, 20]	[20, 40]	[40, 60]	[60, 80]	[80, 100]	[0, 100]
C_{17}	[0, 4]	[4, 7]	[7, 10]	[10, 13]	[13, 15]	[0, 15]
C_{18}	[0, 20]	[20, 40]	[40, 60]	[60, 80]	[80, 100]	[0, 100]
C_{19}	[0, 20]	[20, 40]	[40, 60]	[60, 80]	[80, 100]	[0, 100]

4.2.2. Calculate correlation degree

$$K_j(V_i) = \begin{cases} -\dfrac{\rho(v_i, v_{ji})}{|v_{ji}|} & v_i \in v_{ji} \\ \dfrac{\rho(v_i, v_{ji})}{\rho(v_i, v_{pi}) - \rho(v_i, v_{ji})} & v_i \notin v_{ji}; \rho(v_i, v_{ji}) \neq 0 \\ -\rho(v_i, v_{ji}) - 1 & v_i \notin v_{ji}; \rho(v_i, v_{ji}) = 0 \end{cases} \quad (4)$$

$$\rho(v_i, v_{ji}) = \left|v_i - \frac{1}{2}(a_{ji} + b_{ji})\right| - \frac{1}{2}(b_{ji} - a_{ji}) \quad (i = 1,2,\cdots,n, j = 1,2,\cdots,k) \quad (5)$$

$$\rho(v_i, v_{pi}) = \left|v_i - \frac{1}{2}(a_{pi} + b_{pi})\right| - \frac{1}{2}(b_{pi} - a_{pi}) \quad (i = 1,2,\cdots,n) \quad (6)$$

Equation (4) represents the correlation between the i-th indicator and the j-th level, **Equation (5)** represents the distance between V_i and the classical domain, and **Equation (6)** represents the distance between V_i and the nodal domain. The larger the value of $K_j(V_i)$, the greater the degree of membership of the indicator relative to this level. If $K_j(V_i) = maxK_j(V_i)$, V_i's evaluation object belongs to level j.

Quantitative indicators were analyzed and calculated based on the actual data of the project, and qualitative indicators were scored by experts based on the project's actual situation. The eigenvalue matrix of each

indicator of the SZ distributed photovoltaic power generation project was obtained as p = (10.97, 25.99, 7.69, 66.7, 1.7, 77, 62, 85, 90, 20.90, 54.80, 0.18, 0.15, 69, 71, 73, 8, 57, 66)^T. Through correlation calculation, the evaluation level of each indicator is p = (average, excellent, excellent, poor, good, good, good, excellent, excellent, excellent, good, good, good, good, good, good, average, average, good)^T.

4.2.3. Calculate the comprehensive correlation and determine the evaluation level

$$K_j(N_0) = \sum_{i=1}^{n} \omega_i K_j(V_i) \quad (i = 1,2,\cdots,n) \tag{7}$$

In **Formula (7)**, ω_i represents the weight of each evaluation index, and $K_j(N_0)$ is the comprehensive correlation degree. If the comprehensive correlation is less than -1, it means that the evaluation index does not belong to this level; if the comprehensive correlation is greater than -1 and less than 0, it means that although the evaluation index does not belong to this level, it tends to reach this level. The closer the value is to 0, the stronger the trend toward this level; if the comprehensive correlation is greater than 0, the evaluation index belongs to this level. The evaluation grade correlation of each indicator and comprehensive benefit at the criterion level based on the indicator weight calculation are shown in **Table 5**.

Table 5. Evaluation grade correlation of each indicator and comprehensive benefit at the criterion level

Evaluation index	Poor	Relatively poor	Average	Good	Excellent	Evaluation results
Economic benefits	-0.2900	-0.2573	-0.2503	-0.2588	0.0062	Excellent
Technical benefits	-0.1510	-0.1416	-0.1229	-0.0688	0.0685	Excellent
Environmental benefits	-0.0569	-0.0492	-0.0337	0.0125	-0.0051	Good
Social benefit	-0.0203	-0.0155	-0.0057	0.0072	-0.0089	Good
Overall benefit	-0.2111	-0.1890	-0.1778	-0.1622	0.0204	Excellent

4.3. Result analysis

Through the above analysis, it is evident that indicators such as the project's investment payback period, job creation, and improvement of the energy structure are rated as "average," while the asset-liability ratio indicator is rated as "relatively poor." This suggests that the project is contributing to the transformation of the energy structure and job creation, but the impact is somewhat modest. The investment recovery time could be improved, and there are certain operational risks.

Simultaneously, the comprehensive benefit evaluation grade is "excellent," indicating that the project demonstrates strong overall benefits. Specifically, the economic and technical benefit evaluation grades are "excellent," highlighting the project's robust economic benefits and advanced, reliable technology. On the other hand, the environmental and social benefit evaluation grades are "good," signaling effective energy conservation, emission reduction, and developmental promotion. However, it is acknowledged that further efforts are still required in these areas.

5. Countermeasures and suggestions for improving efficiency

5.1. Promote technological innovation and reduce project costs

To ensure that projects yield optimal comprehensive benefits, it is crucial to promote innovation in photovoltaic technology and exercise stringent cost control. This involves enhancing technological innovation in photovoltaic

power generation, developing photovoltaic modules with higher conversion efficiency, improving the overall efficiency of photovoltaic systems, and gradually achieving the integration of photovoltaic power generation projects with buildings. Such measures will help reduce installation costs and enhance comprehensive benefits by optimizing the layout of photovoltaic components and making informed choices in inverter selection [7].

5.2. Strengthen policy support and expand market presence

In-depth investigations into the operational status of existing distributed photovoltaic power generation projects are essential. A comprehensive understanding of the benefits and challenges arising from the current promotion of these projects should inform an analysis of existing policies. The introduction of Chinese "dual carbon goal" policies can further bolster the healthy development of distributed photovoltaic power generation projects. Additionally, reinforcing market research on these projects will enable a nuanced grasp of policy orientation, market demand, and competition dynamics. Armed with this insight, targeted market expansion strategies can be formulated, and directional adjustments made [8].

5.3. Strengthen the implementation of responsibilities and promote win-win development

Promoting photovoltaic power generation enterprises to align with national policies and guidelines while actively embracing social responsibilities is essential. This involves heightened exploration and research effects, increased investment in research and development, and the ongoing enhancement of China's photovoltaic industry competitiveness [9]. Encouraging the widespread application of clean energy contributes to better environmental and social benefits, fostering the stable development of distributed photovoltaic power generation projects. This, in turn, provides robust support for China's aspirations to build a clean, low-carbon, safe, and efficient energy system, expediting the nation's energy transformation.

6. Conclusion

This article has established a comprehensive benefit evaluation index system focusing on economic, technical, environmental, and social aspects of distributed photovoltaic power generation projects. Through the evaluation and analysis of the SZ distributed photovoltaic power generation project, it is evident that such projects contribute significantly to economic, environmental, and social benefits, aligning with China's green and low-carbon development strategies.

To further advance the field, there is a need to reinforce the comprehensive benefit evaluation indicators, methods, and levels within distributed photovoltaic power generation projects. Research efforts should concentrate on identifying challenges in the current promotion process, offering valuable insights, and presenting targeted countermeasures and suggestions for the promotion and development of distributed photovoltaic power generation projects. This approach is crucial for fostering sustainable progress and supporting China's endeavors in implementing green and low-carbon development strategies.

Disclosure statement

The author declares no conflict of interest.

References

[1] Zhang T, 2019, Research on Comprehensive Benefit Evaluation of Distributed Photovoltaic Power Generation

Projects. Finance and Management, 2019(7): 41–43.

[2] Yan J, Zhang X, Song H, 2021, Comprehensive Analysis of Investment Benefits of Distributed Photovoltaic Power Generation Projects. Modern Marketing, 2021(31): 98–99.

[3] Li J, 2021, Discussion on the Economic Benefits of Distributed Photovoltaic Power Generation. Qianwei, 2021(23): 19–21.

[4] Xu J, Li K, 2022, Research on Distributed Photovoltaic Power Generation Development and Project Economics. China Science and Technology Journal Database Industry A, 2022(8): 48–53.

[5] Hu Y, Li X, Kang X, et al., 2021, Assessment of Rooftop Distributed Photovoltaic Power Generation and Economic Benefits in Inner Mongolia. Arid Area Resources and Environment, 35(12): 66–72.

[6] Qi Z, Wei C, 2022, Research on Full-Process Engineering Consulting Maturity Evaluation Based on AHP-Extension Measurement Model. Journal of Engineering Management, 36(2): 35–40.

[7] Li Y, Yang W, Ling Z, et al., 2022, Research Progress in Building Solar Energy Utilization Technology. Refrigeration and Air Conditioning, 22(8): 92–97.

[8] Qiao Y, 2021, Benefit Analysis and Promotion Measures of Distributed Photovoltaic Power Generation Projects. Chinese Market, 2021(36): 72–73 + 79.

[9] Feng Z, Jia Z, 2023, Thoughts on Developing the Photovoltaic Industry Chain in Shanxi Province. Modern Industrial Economy and Informatization, 13(11): 46–48.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Crossing the Achilles Heel of Algorithms: Identifying the Developmental Dilemma of Artificial Intelligence-Assisted Judicial Decision-Making

Kexin Chen*

School of Law, Macau University of Science and Technology, Macau, China

***Corresponding author:** Kexin Chen, Delores-chen@foxmail.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In the developmental dilemma of artificial intelligence (AI)-assisted judicial decision-making, the technical architecture of AI determines its inherent lack of transparency and interpretability, which is challenging to fundamentally improve. This can be considered a true challenge in the realm of AI-assisted judicial decision-making. By examining the court's acceptance, integration, and trade-offs of AI technology embedded in the judicial field, the exploration of potential conflicts, interactions, and even mutual shaping between the two will not only reshape their conceptual connotations and intellectual boundaries but also strengthen the cognition and re-interpretation of the basic principles and core values of the judicial trial system.

Keywords: Artificial intelligence; Automated decision-making; Algorithmic law system; Due process; Algorithmic justice

Online publication: January 23, 2024

1. Introduction

In recent years, artificial intelligence has become a pivotal component of judicial decision-making on a global scale. Leveraging advancements in deep learning, machine learning, natural language processing, and other technological innovations, it has propelled the informatization of legal affairs beyond simple electronic retrieval to encompass more intelligent functions. These include class case push, deviation warning, danger assessment, sentencing assistance, sentence prediction, and more.

Despite the significant strides made, the application of artificial intelligence (AI)-assisted judicial decision-making evokes both optimism and pessimism. Among the challenges faced by current AI-assisted judicial decision-making systems in practical use, one that stands out as an essentially insurmountable "Achilles' heel" is the algorithmic black box. This paper seeks to delve deep into the exploration of this issue.

2. Legal black boxes and transparency

The 2016 case of State v. Loomis drew widespread attention to the controversy surrounding risk assessment

algorithms in the United States. In this instance, the court imposed a severe sentence based on COMPAS' assessment that the defendant posed a high risk. Subsequently, the defendant moved to reopen the sentencing process, contending that the court's utilization of the COMPAS system violated constitutionally guaranteed due process rights [1].

One of the defendant's arguments was that the COMPAS system, protected by trade secrets, could not be disclosed for an assessment of its scientific accuracy [1]. The most straightforward solution for the "legal black box" issue would be to mandate the judiciary to disclose the original algorithm code when contracting an enterprise to develop an intelligent court system. However, considering that the trade secret clause safeguards developers' rights in highly specialized fields and encourages research and development in emerging technologies, mandatory disclosure might undermine the legitimate interests of algorithm developers. If mandatory disclosure poses a threat to these rights and interests, it should be carried out within reasonable limits, even for due process protection.

Kroll and his colleagues suggested, from a technical standpoint, that the operational steps of algorithmic decision-making can be recorded through specific technical means. For example, zero-knowledge proofs can enable decision-makers to prove that system decisions meet specific requirements without revealing all the contents of cryptographic commitments. Alternatively, cryptographic commitments can be adapted to prove algorithmic decision-making's compliance with specific requirements. In this context, a third party may keep sealed files or numbers to ensure a review of decision-making in a non-public manner, or software verification can mathematically ensure the existence of invariants in the system to guarantee a degree of accountability for the algorithm [2].

In contrast, Contini took a regulatory perspective, suggesting that additional expert committees could be established to review the compliance of algorithmic systems with legal norms, ensuring "qualified transparency" [3]. However, even if algorithms are entirely "transparent," they cannot meet the requirement of accountability based on the understanding of the intelligent court system operation logic. The "technological black box" inherently gives rise to a transparency paradox. In other words, algorithmic transparency does not equate to comprehensibility, let alone accountability [4].

3. Technical black boxes and interpretability

In the Loomis decision, the discussion regarding whether the court's use of the COMPAS system violated due process rights only brought attention to the "legal black box" aspect, neglecting the more pressing issue of the "technological black box" [1]. Indeed, the lack of interpretability stemming from the "technical black box" is the most challenging flaw in fundamentally improving algorithmic systems.

The most direct solution to address the interpretability gap in AI decision-making is to disclose the original program code, allowing for the dissection of the algorithm's logic. However, this approach holds little meaning for non-specialists who cannot glean valuable information by merely reading the code [5]. Moreover, it fails to alleviate the concerns of those affected by the decision regarding the algorithm's result reliability [2]. Additionally, due to the unique nature of deep learning and neural network algorithms, professionals in the field remain in the dark about how the system adjusts the weights of various factors and its operational logic [6]. Furthermore, as various reference factors within the algorithm may dynamically adjust at any time, there is no guarantee of result predictability, let alone repeatability [5].

The issue of algorithmic interpretability cannot be adequately addressed at the current level of existing technologies and specifications. On the technical front, explainable AI technology has gradually developed

to alleviate this problem. Gleicher advocated intentionally constructing "interpretable models" in the future to intervene in the algorithmic selection process by controlling variables [7]. However, this approach faces an algorithmic design paradox: the intelligence and accuracy of algorithms cannot seamlessly coexist with interpretability. Only simple algorithms can maintain a high level of interpretability, while the interpretability of complex algorithms always falls short.

On the legal norms front, the General Data Protection Regulation (GDPR) introduced by the European Union (EU) in 2018 provides a more comprehensive norm for algorithm-related responsibility. Although it does not explicitly outline the "right to interpret" automated decision-making. Articles 13-15 and 22 of the GDPR address the issue. Article 13-14 stipulates that the data subject has the right to know the automated decision-making process and information related to expected results. However, the notification obligation is outlined before data collection, not after the decision results are formed, making it challenging to use as the basis of the right to explanation after the algorithm's decision-making results are generated.

Some scholars further argue that the right of access in Article 15 may serve as the jurisprudential basis for the right of interpretation. Still, since the point in time for exercising this power is not explicitly stated, it can be understood, at most, as the right of interpretation for the system's functionality rather than the basis for requesting interpretation in automated decision-making. Article 22, addressing a party's right to reject fully automated decision-making with a significant impact on its rights, is not essentially an affirmative claim but more akin to a passive right of defense. The "technological black box" has yet to provide a proper response to the interpretability problem it gives rise to.

4. How to address the impact of algorithmic black boxes on due process

The primary question that requires attention is whether the lack of transparency and interpretability resulting from the "algorithmic black box" is an issue that cannot be overlooked. Techno-optimists, such as Eugene Volokh, argued that when exploring the appropriateness of AI algorithms in judicial decisions, the focus should be on the outcomes rather than understanding the decision-making process. However, this view has faced opposition from some scholars. They contended that if the emphasis is solely on the results of smart court outputs rather than the process, it not only infringes upon the parties' right to contest adverse evidence but, more significantly, undermines the court's potential role in dispute resolution. The court plays a crucial role not only in soothing people's concerns but also in repairing relationships between parties and fostering reconciliation in society.

Given the critical importance of transparency and interpretability, and acknowledging the difficulty of resolving the issue at the technological source level, it might be feasible to address it through rights remedies. Firstly, the scope of due process rights violations resulting from AI-assisted judicial decision-making must be defined. According to Loomis's defense on appeal, due process rights were violated because COMPAS, safeguarded by trade secrets, prevented a challenge to the scientific validity of the risk assessment. In other words, the court encroached on its "right to interpret" based on State v. Skaff, established by the court. This right entails the defendant's ability to review the accuracy of the judgment and factors influencing it, including understanding the risk assessment algorithm and how danger assessment results were generated. AI fundamentally and directly affects the litigants' rights, especially their right to effective defense.

In the future, in criminal judgments, courts may be required to explicitly state whether they have considered decision-making aided by the risk assessment system and disclose specific risk factors analyzed by the system. For instance, the "Sentencing Trend Suggestion System" developed by the Judicial Yuan of Taiwan publicizes reference factors on its website. Taking Article 221 of the Criminal Law of Taiwan on the crime of compulsory

sexual intercourse as an example, the system lists selected reference factors such as the perpetrator's age at the time of the act, previous cases of obstruction of sexual autonomy, confession consistency, and the relationship with the victim. Such transparency measures enhance the understanding and accountability of AI-assisted decision-making in the judicial context.

5. Conclusion

Given the inevitability of embracing the tide of technological development, confronting the impact of AI algorithms on the existing judicial system demands a nuanced approach. Treating AI technology is not a zero-sum game of weighing "advantages" against "shortcomings." Instead, the impact of AI technology on core values and the basic principles of the current judicial system requires careful examination and contemplation.

For instance, the lack of transparency and interpretability resulting from the algorithmic black box can significantly impact due process. Therefore, examinations should extend to understanding how the current judicial system accepts, integrates, and negotiates the trade-offs of AI technology. The goal is to ensure that human judges maintain trial independence and that the rights and obligations of the parties are effectively safeguarded. Simultaneously, science and technology may be harnessed to enhance the quality and efficiency of judicial proceedings, striving for a delicate balance and self-consistency within the system.

Disclosure statement

The author declares no conflict of interest.

References

[1] Liu H-W, Lin C-F, Chen Y-J, 2019, Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability. International Journal of Law and Information Technology, 27(2): 122–141.

[2] Kroll JA, Huey J, Barocas S, et al., 2017, Accountable Algorithms. University of Pennsylvania Law Review, 165(3): 633–705.

[3] Contini F, 2020, Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings. Law, Technology and Humans, 2(1): 4-18. https://doi.org/10.5204/lthj.v2i1.1478

[4] Schönberger D, 2019, Artificial Intelligence in Healthcare: A Critical Analysis of the Legal and Ethical Implications. International Journal of Law and Information Technology, 27(2): 171–203. https://doi.org/10.1093/ijlit/eaz004

[5] McGregor L, Murray D, Ng V, 2019, International Human Rights Law as a Framework for Algorithmic Accountability. International and Comparative Law Quarterly, 68(2): 309 – 343. https://doi.org/10.1017/S0020589319000046

[6] Ananny M, Crawford K, 2016, Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. New Media & Society, 20(3): 973–989. https://doi.org/10.1177/1461444816676645

[7] Gleicher M, 2016, A Framework for Considering Comprehensibility in Modeling. Big Data, 4(2): 75–88. http://doi.org/10.1089/big.201$^{6.0007}$

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Simulation-Based Investigation of Crack Formation in Forged MB15 Magnesium Alloy

Lulu Song[1], Dongmei Liu[1], Jie Yuan[1], Ruihuan Tian[1], Liying Jin[2]*

[1]Metrological Physics and Chemistry Center, AVIC Aerospace Life-Support Industries, Ltd., Xiangyang 441003, Hubei Province, China
[2]Xiangyang Key Laboratory of Intelligent Manufacturing and Machine Vision, Hubei University of Arts and Science, Xiangyang 441053, Hubei Province, China

***Corresponding author:** Liying Jin, JLY@hbuas.edu.cn

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In the process of riveting an MB15 forging die, cracks were discovered emerging along the longitudinal direction and near the riveting hole. Through fracture analysis, microscopic observation, energy spectrum analysis, metallographic examination, and hardness test, the properties and causes of the cracks are discussed. The results indicate that the cracking type is intergranular brittle cracking, occurring during the forging stage. Furthermore, the recrystallization at the crack site is found to be incomplete, which is attributed to the low deformation temperature.

Keywords: MB15; Forging die; Crack; Deformation temperature

Online publication: January 23, 2024

1. Introduction

MB15 is a high-strength deformed magnesium alloy belonging to the magnesium-zinc-zirconium series, capable of being strengthened through heat treatment. The process properties of this alloy, with lower plasticity compared to the medium-strength MB2, MB3, and MB8 alloys, resulted in a limited production range, mainly focused on extruded products, forgings, and forging dies. The alloy finds primary applications in the production of aircraft girders, rocker arms in the operating system, supports, and other stress components [1].

During the assembly and riveting of a batch of MB15 deformed magnesium alloy forgings, cracks were discovered along the longitudinal axis and near the riveting holes. The production process involves heating the bar die, forging and pressing the billet, milling the billet die, heating the billet die, two forging forming, grinding the flash, heat treatment, surface treatment, lettering, machine addition, surface oxidation, painting, and assembly. Through macroscopic observation, microscopic observation, energy spectrum analysis, metallographic inspection, and hardness test of the crack fracture, the properties and causes of the crack are discussed.

2. Test process

2.1. Macroscopic observation

After undergoing chemical oxidation and paint treatment, the forging surface appears black. Cracks are distributed parallel to the forging flow line, with some penetrating the entire thickness from the outer wall, while others do not. These cracks exhibit a flat and straight nature, and local positions reveal the presence of multiple cracks, as depicted in **Figure 1**.

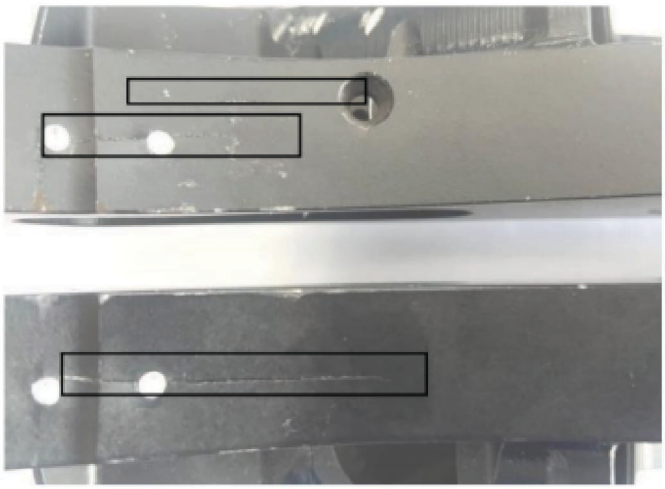


Figure 1. Macroscopic morphology of cracks

For macroscopic observation, the crack location was manually opened. Cracks extend from the outer wall towards the interior, resulting in relatively level fractures. The original crack area appears dark gray, and the local crack area exhibits a yellowish-brown tint with a noticeable small step characteristic. The artificially interrupted fracture appears silver-gray, as illustrated in **Figure 2**. The original cracked area and artificially interrupted area were observed respectively under stereomicroscope, and the step characteristics and reflective features were visible in the original cracked area, as shown in **Figure 2**. The fracture in the artificially interrupted area was fibrous, as shown in **Figure 2**.



Figure 2. Macroscopic morphology of the fracture

2.2. Microscopic observation

The crack fracture was examined under a scanning electron microscope (SEM). In the initial crack zone, the fractures appeared flush, covered with a substantial amount of oxides. Local, clear intergranular characteristics were observed, and no metallurgical defects were present, as illustrated in **Figure 3(a)**. During the middle stage of expansion, the fracture exhibited evenness, displaying layered cracking characteristics. Upon magnification,

the fracture was found to be intergranular, with noticeable oxide adhesion on the crystal surface, as depicted in **Figure 3(b)**. The characteristics of fractures at the end stage of the expansion were similar to the middle stage, showing intergranular with the crystal surface exhibiting oxidation, as shown in **Figure 3(c)**. Fractures resulting from artificial interruption displayed cleavage-like features, as depicted in **Figure 3(d)**.

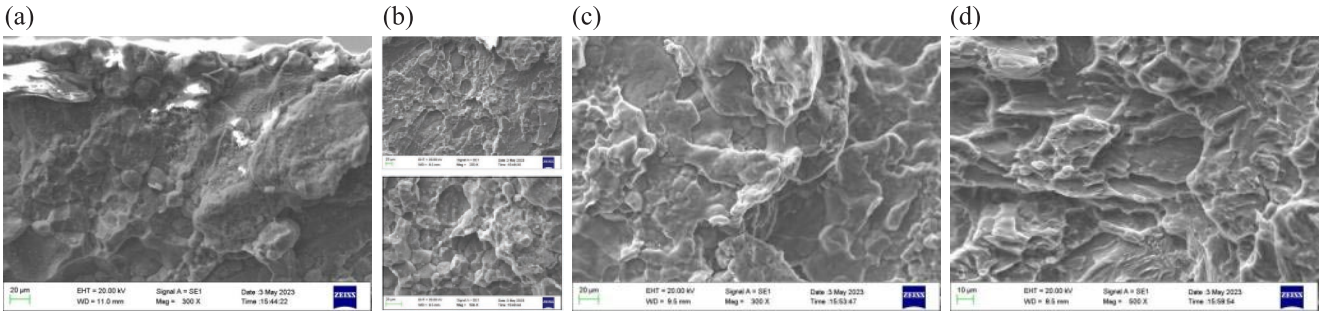
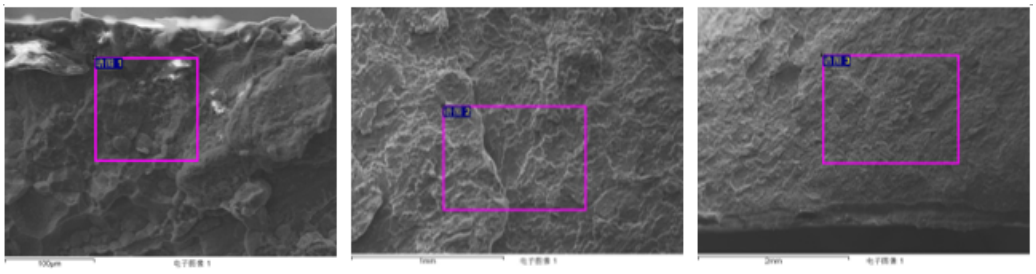


Figure 3. Microscopic observation of the crack fracture. **(a)** Initial cracking zone; **(b)** Middle stage of expansion; **(c)** End stage of expansion; **(d)** Artificial interruption of the fracture

2.3. Energy spectrum analysis

The crack's source region, expansion region, and artificially induced fracture region were analyzed separately. In the crack source region, additional elements such as O, F, Na, and Cr were identified. Elements such as F, Na, and Cr originated from chemical oxidation, with the surface crack fracturing during this process, as illustrated in **Figure 4(a)**. The expansion zone exhibited a higher concentration of O elements, while the types and contents of the main alloy elements essentially met the requirements of the MB15 alloy, as shown in **Figure 4(b)**. In the artificially induced fracture, the O element content was significantly reduced due to natural oxidation, yet the types and contents of the primary alloy elements met the specifications of the MB15 alloy, as depicted in **Figure 4(c)**.



(a)

Elements	Weight Percent	Atoms Percent
O K	31.08	38.60
F K	7.51	7.85
Na K	2.25	1.94
Mg K	27.90	22.80
Si K	0.92	0.65
S K	0.40	0.25
Cl K	1.01	0.57
K K	0.67	0.34
Ca K	0.58	0.29
Cr K	10.54	4.03
Zn K	4.19	1.27
C	12.94	21.40
Total	100.00	

(b)

Elements	Weight Percent	Atoms Percent
O K	15.58	21.53
Mg K	72.04	65.52
Zn K	6.25	2.11
Zr L	0.28	0.07
C	5.85	10.77
Total	100.00	

(c)

Elements	Weight Percent	Atoms Percent
O K	2.05	3.21
Mg K	89.74	92.37
Zn K	7.13	2.73
Zr L	0.31	0.08
C	0.77	1.61
Gross amount	100.00	

Figure 4. Energy spectrum analysis of the crack. **(a)** Source area; **(b)** Expansion area; **(c)** Artificial interruption area

2.4. Metallographic analysis

Metallographic structure detection was performed at both the crack location and the intact location. It was observed that the cracks primarily propagated along the crystals, demonstrating good alignment between the two sides of the cracks. Simultaneously, a substantial number of elongated deformed grains and deformation bands were evident, with occasional occurrences of equiaxed crystals in local areas, as illustrated in **Figure 5(a)**. During the expansion stage, secondary cracks appeared on both sides of the main crack, extending along the crystal structure, as depicted in **Figure 5(b)**. No signs of overheating or overburning were observed in the structure. The normal position's metallographic structure morphology is presented in **Figure 5(c)**.

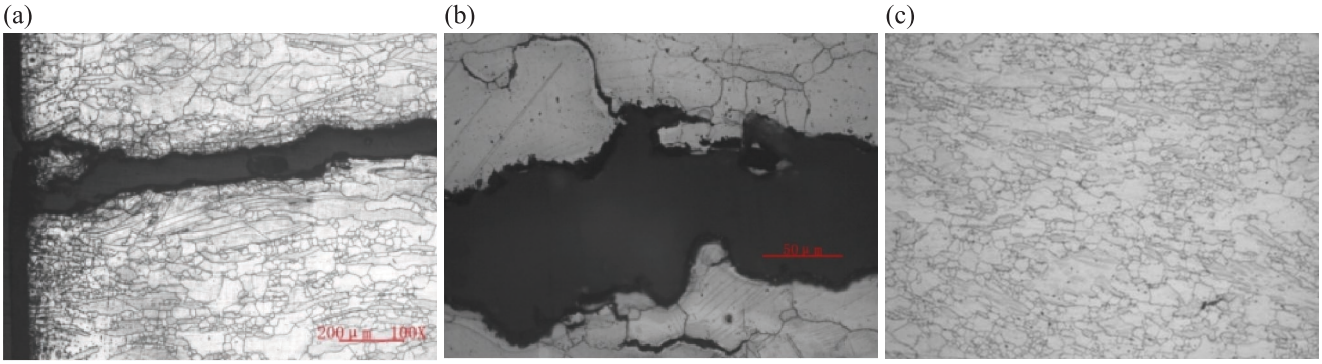


Figure 5. Metallographic analysis. **(a)** Initial cracking zone; **(b)** Expansion area; **(c)** Metallographic structure at normal position

2.5. Hardness inspection

The sample was cut near the crack for the Brinell hardness test, and the results are presented in **Table 1**. The hardness exhibited uniformity, with an average hardness of HB86, meeting the design requirement of HB260.

Table 1. Brinell hardness test results (HB2.5/62.5)

1	2	3	Average value
85	86	87	86

3. Analysis and discussion

The forging cracks are oriented parallel to the forging flow line and propagate from the external to the internal regions. They exhibited evenness, devoid of metallurgical defects, with the original crack area appearing dark gray, locally yellowish-brown, and the artificially interrupted fracture appearing silver-gray. The initial cracking zone of the crack was covered with more oxidation products, displaying distinct layered cracking and intergranular fracture characteristics. The expansion zone fracture is intergranular, with visible oxidation on the crystal surface, while the artificial interrupted fracture exhibits quasi-cleavage characteristics.

In summary, it is deduced that the forging cracking is intergranular and brittle. The structure near the crack consists of elongated deformed grains and a few fine equiaxed crystals, with a visible deformation zone, indicating a low degree of recrystallization. The main factors influencing recrystallization include the original material structure, deformation rate, and deformation temperature. The degree of recrystallization away from the crack is higher than at the crack location, suggesting that the main factors affecting recrystallization at the crack location may be a rapid deformation rate or low deformation temperature during forging. However, a rapid forging deformation rate may increase the number of crystal twinning, resulting in a higher hardness value

in positions with more twins [2]. Oxidation products are observed in the initial zone of crack fracture, indicating that the crack occurred before surface treatment. Additionally, only oxidation characteristics are evident inside the crack fracture, supporting the belief that the crack originated during the forging stage.

The deformation process of magnesium alloy primarily involves dislocation slip and twinning. Dislocation slip is hindered by grain boundaries, second phases, twin boundaries, etc., resulting in increased dislocation density within the magnesium alloy. At high temperatures, the dislocation with high density can induce recrystallization, resulting in grain refinement of magnesium alloy. Fine grains are easy to slip along grain boundaries and coordinate deformation. Dislocation at high temperatures makes it easier to slip or climb with less resistance [3]. Therefore, the magnesium alloy is conducive to promoting deformation at high temperatures. When the forging temperature is relatively low, the degree of dynamic recrystallization is weak, resulting in incomplete dynamic recrystallization, un-recrystallized structure, and hindered dislocation climbing. Subsequent deformation processes result in dislocation blockage at the grain boundary, initiating microcracks. The connection of microcracks ultimately leads to forging cracks along the crystal. Thus, the primary cause of cracking is the low forging temperature.

4. Simulation analysis

A simulation analysis was conducted on the forging mold with the mold structure detailed in **Table 2**. During forging, the blank is positioned in the lower mold, and upon the initiation of the upper mold's downward movement, the blank takes shape, reaching the desired displacement. The simulation analysis employed finite element software, and the parameters are outlined in **Table 3**. The design requirements specify that the forging process must be devoid of defects such as incomplete filling and folding, with the forging tonnage not exceeding 6000 T (in line with the equipment capacity of 6000 T).

Table 2. Mold structure and forming tonnage diagram

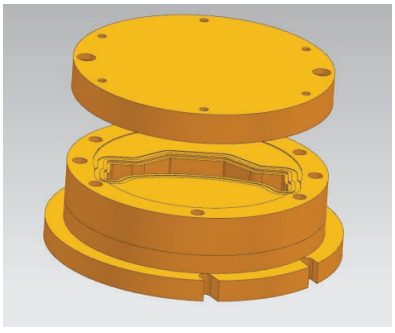
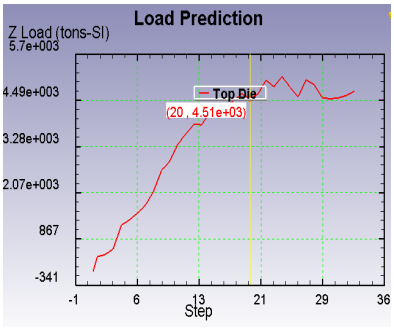
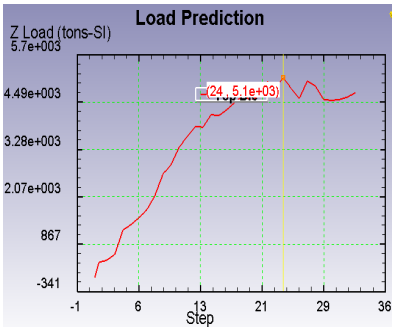
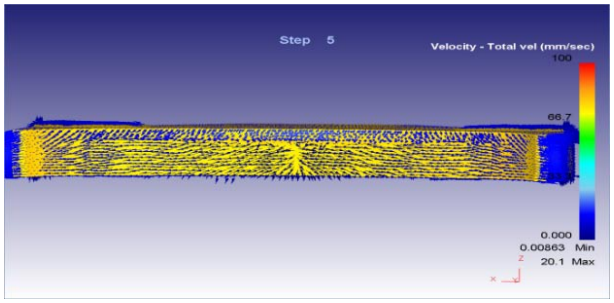
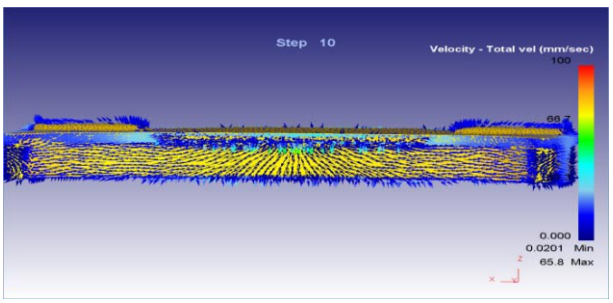
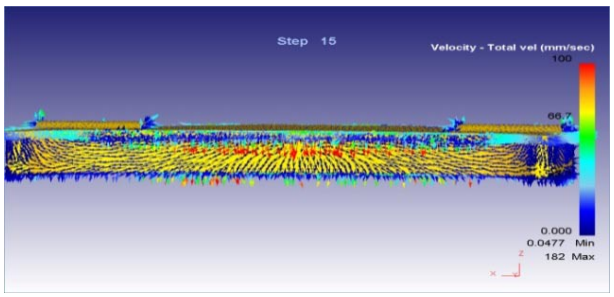
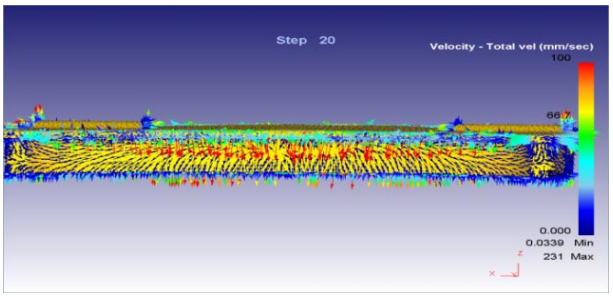
Mold structure diagram	Forming tonnage	
		

Table 3. Simulation parameter diagram

Project	Temperature (°C)	Number of grids	Follow mold displacement	Step count
Blank	340	106,440		
Upper mold	350	115,020	0.35 mm	20
Lower mold	350	102,760		

Upon physical inspection (refer to **Table 4**), it is evident that during secondary forging, the material flows to both sides, and the forging fibers follow a horizontal direction. The low temperature of the blank induces the initiation of forging cracks along the fiber direction, aligning with the results observed in the metallographic analysis.

Table 4. Blank flow diagram

5. Conclusions

The cracking property of the part is intergranular brittle cracking. The recrystallization structure of the cracking part is not complete, which is related to the low forging temperature in the forging process.

Authors contribution

Funding acquisition, **Liying Jin**; Supervision, **Dongmei Liu**, **Jie Yuan**, **Ruihuan Tian**; Validation, **Lulu Song**.

Funding

Key R&D Plan Projects in Hubei Province (Grant No. 2021BID001); the Research on Multiple Regression and Fitting Technology of Simulation Data for Dynamic Umbrella Opening of Lifesaving Umbrella (Grant No. HX2021157)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Editorial Committee of China Aviation Materials 2002, China Aviation Materials Handbook, Vol. 3, Aluminum Alloy,

Magnesium Alloy. 2nd Edition, China Standards Press, Beijing.

[2] Dai L, Zhang R, Liu X, 2007, The Effect on Structure and Mechanics Property of MB15 Magnesium Alloy Deformation. Inner Mongolia Petrochemical Industry, 2007(11): 59–60.

[3] Xiang J, Wang C, Chen K, et al., 2019, Surface Modification and Corrosion Resistance of MB15 Magnesium Alloy for Automobiles. Heat Treatment of Metals, 44(12): 186–192.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Attribute Reduction of Neighborhood Rough Set Based on Discernment

Biqing Wang*

School of Mathematics and Computer, Tongling University, Tongling 244000, Anhui Province, China

***Corresponding author:** Biqing Wang, wbq@tlu.edu.cn

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: For neighborhood rough set attribute reduction algorithms based on dependency degree, a neighborhood computation method incorporating attribute weight values and a neighborhood rough set attribute reduction algorithm using discernment as the heuristic information was proposed. The reduction algorithm comprehensively considers the dependency degree and neighborhood granulation degree of attributes, allowing for a more accurate measurement of the importance degrees of attributes. Example analyses and experimental results demonstrate the feasibility and effectiveness of the algorithm.

Keywords: Neighborhood rough set; Attribute reduction; Discernment; Algorithm

Online publication: January 23, 2024

1. Introduction

As Artificial Intelligence Generated Content (AIGC) becomes an increasingly prominent research focus, the analysis and processing of data play an even more crucial role. Rough set theory can extract essential features from a dataset while ensuring the correct classification rate [1]. The traditional Pawlak rough set model is suitable only for handling complete symbolic data. Numerical data, prevalent in real applications, cannot be processed directly; hence, data need to be discretized before reduction. However, the discretization process often results in information loss [2]. The neighborhood rough set model allows for the direct reduction of the attributes in continuous-type data, eliminating the information loss associated with discretization [3].

In the context of neighborhood rough set models, the determination of the neighborhood is of paramount importance. Typically, when calculating the neighborhood, the same weight value is assigned to all attributes. However, in practical scenarios, attributes often exhibit significant differences in the data, leading to varying degrees of importance [4]. Assigning identical weight values to different attributes neglects these differences and may result in suboptimal attribute reduction [5]. Therefore, this paper proposes to set different weight values for various conditional attributes based on experience and experiments when calculating the neighborhood. This approach addresses the challenge of the original neighborhood rough set's difficulty in distinguishing the importance of different conditional attributes.

Furthermore, existing neighborhood rough set attribute reduction algorithms measure attribute importance using dependency-based attribute importance degree as heuristic information [6]. Dependency-based attribute importance degree is defined by the difference between dependency degrees before and after the inclusion of the attribute [7]. However, this definition does not account for the neighborhood granulation degree of the attribute. In reality, the neighborhood granulation degree also reflects an attribute's ability to discriminate between objects. Two attributes causing the same dependency change cannot be considered equally important without considering their discriminative abilities through the neighborhood granulation degree. Hence, this paper introduces the concept of neighborhood granulation degree and constructs an improved evaluation metric of attribute importance, namely discernment, by combining dependency degree and neighborhood granulation degree. Based on this, a heuristic attribute reduction algorithm for neighborhood rough sets is designed, using discernment as the heuristic information. The proposed algorithm's feasibility and effectiveness are then verified through examples and experiments.

2. Preliminaries

The relevant definitions of neighborhood rough set are presented below [8,9]:

Definition 1: Given a M-dimensional real number space Ω, $\Delta: R^M \times R^M \to R$, Δ is called a metric on R^M. For any $y_1, y_2, y_3 \in R^M$, if Δ meets the following conditions:

(1) $\Delta(y_1, y_2) \geq 0$, $y_1 = y_2$, $\Delta(y_1, y_2) = 0$;

(2) $\Delta(y_1, y_2) = \Delta(y_2, y_1)$;

(3) $\Delta(y_1, y_3) \leq \Delta(y_1, y_2) + \Delta(y_2, y_3)$;

$\langle \Omega, \Delta \rangle$ is called metric space. For any two points in space $y_i = (y_{i1}, y_{i2}, \ldots, y_{iM})$ and $y_j = (y_{j1}, y_{j2}, \ldots, y_{jM})$, the distance between y_i and y_j is often expressed in terms of the P-paradigm as:

$$\Delta_P^M(y_i, y_j) = \left(\sum_{k=1}^{M} |y_{ik} - y_{jk}|^P\right)^{1/P} \quad (1)$$

Definition 2: Given a nonempty finite set on the space of real numbers $U = (y_1, y_2, \ldots, y_n)$, for any object y_i in U, its δ neighborhood is defined as:

$\delta(y_i) = \{y | y \in U, \Delta(y, y_i) \leq \delta\}$ (2)

where $\delta \geq 0$ is the neighborhood radius, and $\delta(y_i)$ is the neighborhood particle of y_i.

Definition 3: Let U be a set of objects, C a set of conditional attributes, and D a set of decision attributes. If C can generate a cluster of neighborhood relations on U, $NDA = \langle U, C, D \rangle$ is called a neighborhood decision-making system.

3. Proposed algorithm

In practice, the importance of various conditional attributes differs. Some attributes determine the primary characteristics of an object, while others play only a minor role in describing the object [10]. To capture this variability in conditional attributes, this paper assigns a weight value to each conditional attribute. This ensures that the calculation of distance and neighborhood reflects the primary characteristics of the object. The specific values of the weights are determined through experiments and experiences. Building upon the preceding definitions, the following definitions are provided:

Definition 4: The P-paradigm distance with attribute weight values is defined as:

$$\Delta_P^{M'}(y_i, y_j) = \left(\sum_{k=1}^{M} \left|\alpha_k y_{ik} - \alpha_k y_{jk}\right|^P\right)^{1/P} \quad (3)$$

where a_k is the weight of the kth attribute and $a_k \in [0, 1]$, $\sum_{k=1}^{M} a_k = 1$.

Definition 5: The δ neighborhood of y_i with attribute weight values is defined as:

$\delta'(y_i) = \{y | y \in U, \Delta_P^{M'}(y, y_i) \leq \delta\}$(4)

Existing dependency-based neighborhood rough set attribute reduction algorithms do not consider the effect of attributes on granulation degree. To solve this problem, the concept of neighborhood granulation degree is introduced below.

Definition 6: Given a neighborhood decision system $NDA = \langle U, C, D \rangle$, for any subset of attributes $Q \subseteq C$, $\delta'(y_i)$ is the neighborhood of y_i on Q. The neighborhood granulation degree of Q is defined as:

$$NGD(Q) = \frac{|\delta_Q'(y_1)|^2}{|U|^3} + \frac{|\delta_Q'(y_2)|^2}{|U|^3} + \ldots + \frac{|\delta_Q'(y_{|U|})|^2}{|U|^3} \quad (5)$$

Neighborhood granulation degree is a measure of the discriminative ability of an attribute set. Obviously, the smaller the radius and the more balanced the size of the neighborhood particles, the stronger the ability of the attribute set to discriminate the object, and the smaller its neighborhood granulation degree. If for all $y_i \in U$, the neighborhood of y_i on Q is $\{y_i\}$, and $NDA(Q)$ obtains the minimum value $1/|U|^2$. If for all $y_i \in U$, the neighborhood of y_i on Q is U, $NDA(Q)$ gets the maximum value of 1. In particular, $NDA(\varnothing) = 0$.

Subsequently, an improved attribute importance metric function is defined based on neighborhood granulation degree, namely discernment, was presented in the following:

Definition 7: Given a neighborhood decision system, $A \subseteq C$, for $\forall \beta \in U$, the discernment of β concerning A-β and D is defined as:

$DIS(\beta, A, D) = \Delta\gamma + \Delta NGD$(6)

where γ is the degree of dependence as defined by rough set theory [9], $\Delta\gamma = \gamma_A(D) - \gamma_{A-B}(D)$, and $\Delta NGD = NGD(A-\beta) - NGD(A)$.

From this definition, it can be seen that the improved attribute importance degree combines the dependency degree and neighborhood granulation degree. It can evaluate the classification ability of attributes more objectively, and it is a more comprehensive and precise attribute importance degree measure function. Now the improved neighborhood rough set attribute reduction algorithm is proposed using discernment as the heuristic function, as shown in **Algorithm 1**.

Algorithm 1: Neighborhood rough set attribute reduction algorithm based on discernment

Input: $NDS = \langle U, C, D \rangle$

Output: A reduction R for neighborhood decision system

$R = \varnothing$;

compute $\gamma_C(D)$;

while $(\gamma_R(D) < \gamma_C(D))$

{for each conditional attribute $\beta \in C-R$, compute $DIS(\beta, A, D)$;

select the β with the largest discernment;

$R = R \cup \{\beta\}$;}

return R.

Obviously, the key of the algorithm is to continually select the conditional attribute with the largest discernment and add it to the reduction set until the dependency of the reduction set equals that of the original conditional attributes set. The time complexity of computing the neighborhoods for all objects is $0(|C||U|)$.

Dependency is computed based on the neighborhood of the object, resulting in a time complexity of $0(|C||U|)$. The time complexity of computing the discernment within a while loop is $0(|C|^2|U|)$. In the worst case, the discernment is computed $|C| + (|C|–1) + (|C|-2) + \ldots +3+2+1 = (|C| + 1)|C|/2$. Therefore, the time complexity of the while loop is $0(|C|^2|U|)$. Thus, the total time complexity of **Algorithm 1** is $0(|C|^4|U|)$.

4. Experimental results

The feasibility of the neighborhood rough set attribute reduction algorithm based on discernment has been verified by the example in the previous section. In this section, an experimental comparison was made between the neighborhood rough set attribute reduction algorithm based on the dependency model (Algorithm A) [11] and the proposed algorithm (Algorithm B) in this paper to further validate the effectiveness of the proposed algorithm. Three datasets – Sonar, Ionosphere, and Vehicle – from the UCI database were used in the experiments. **Table 1** presents specific information about the experimental datasets.

Table 1. Data description

Data set	Samples	$\|C\|$	Classes
Sonar	208	60	2
Ionosphere	351	34	2
Vehicle	846	18	4

The hardware and software environment for the experiment is as follows: Processor: Intel i7-7500U; Memory: 8 GB; Operating System: 64-bit Windows 10; Software: Python 3.9. Numerical attributes in the dataset often have different units of measure. This variability may affect the reduction results. For this reason, the values of all numerical conditional attributes in the dataset have been normalized to the interval [0, 1] before attribute reduction. For the selection of the neighborhood radius, the recommended value $\delta = 0.2$ was used [8]. Meanwhile, to measure the classification ability of the reduction attribute set, the classification accuracy of the reduction attribute set is calculated using an SVM classifier. The experimental results are shown in **Table 2**.

Table 2. Experimental results

Data set	Algorithm A		Algorithm B	
	$\|R\|$	Accuracy (%)	$\|R\|$	Accuracy (%)
Sonar	18	72.5	7	75.3
Ionosphere	8	89.6	6	94.1
Vehicle	10	71.0	7	76.4

As observed in **Table 2**, the number of attributes in the reduction results of the attribute reduction algorithm proposed in this paper is less than that of the neighborhood rough set attribute reduction algorithm based on the dependency model for all three datasets. Moreover, the classification accuracies of the obtained reduction results are all higher than the latter, indicating that the proposed algorithm in this paper is effective.

5. Retrospect and prospect

In this paper, a definition of discernment in conjunction with neighborhood granulation degree, building

upon existing neighborhood rough set attribute reduction algorithms based on the dependency model, was provided. The basic properties of this improved attribute importance degree were discussed. Subsequently, a neighborhood rough set attribute reduction algorithm that leverages discernment's ability to discriminate between objects was proposed, and a detailed description of the algorithm's steps was provided. The feasibility of the proposed algorithm was confirmed through an illustrative example. Finally, the proposed algorithm and the traditional algorithms were compared and analyzed using three datasets from the UCI database. The experimental results demonstrated that the proposed algorithm achieved higher classification accuracy. In the next step, the algorithm could be applied to construct a knowledge graph for manufacturing production lines.

Funding

Anhui Provincial University Research Project (Project Number: 2023AH051659); Tongling University Talent Research Initiation Fund Project (Project Number: 2022tlxyrc31); Tongling University School-Level Scientific Research Project (Project Number: 2021tlxytwh05); Tongling University Horizontal Project (Project Number: 2023tlxyxdz237)

Disclosure statement

The author declares no conflict of interest.

References

[1] Pawlak Z, 1982, Rough sets. International Journal of Computer & Information Science, 11: 341–356. https://doi.org/10.1007/BF01001956

[2] Deng Z-X, Zheng Z-L, Deng D-Y, 2021, F-Neighborhood Rough Sets and Its Reduction. Acta Automatica Sinica, 47(3): 695−705. https://doi.org/10.16383/j.aas.c180556

[3] Ji J-Z, Song X-N, Yang C-C, 2020, Feature Reduction of Neighborhood Rough Set Based on Fish Swarm Algorithm in Brain Functional Connectivity. Journal of Zhejiang University (Engineering Science), 54(11): 2247–2257. https://doi.org/10.3785/j.issn.1008-973X.2020.11.020

[4] Chen T, Liu J, Zhu F, et al., 2018, A Novel Multi-Radius Neighborhood Rough Set Weighted Feature Extraction Method for Remote Sensing Image Classification. Geomatics and Information Science of Wuhan University, 43(2): 311–317. https://doi.org/10.13203/j.whugis20150290

[5] Wu Y, Zhao R, Jin W, 2019, Fault Feature Extraction of Rotating Machinery Based on EWT and a Weighted Multi Neighborhood Rough Set. Journal of Vibration and Shock, 38(24): 235–242.

[6] Wang H, Wang S, Yao J, et al., 2019, Tag Identification Rate Prediction Based on Neighborhood Rough Set and Support Vector Machine. Computer Integrated Manufacturing Systems, 25(12): 3170–3180.

[7] Zhang W, Wu W, Liang J, et al., 2001, Rough Set Theory and Method. Science Press, Beijing.

[8] Hu Q-H, Yu D-R, Xie Z-X, 2008, Numerical Attribute Reduction Based on Neighborhood Granulation and Rough Approximation. Journal of Software, 19(3): 640–649.

[9] Liu X, Wang X-L, 2021, Value Segmentation of Airline Customers Based on K-means and Neighborhood Rough Set. Operations Research and Management Science, 30(3): 104–111.

[10] Chen Y-M, Zhu Q-X, Zeng Z-Q, et al., 2018, Gene Selection Method Based on Neighborhood Rough Sets and Fish Swarm Intelligence. Journal of University of Electronic Science and Technology of China, 47(1): 99–104.

[11] Yong L, Huang W, Jiang Y, et al., 2014, Quick Attribute Reduct Algorithm for Neighborhood Rough Set Model. Information Sciences, 271: 65–81. https://doi.org/10.1016/j.ins.2014.02.093

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Synthesis of Ionic and Non-Ionic Modified Waterborne Polyurethane Curing Agent

Ying Qiao[1], Xu-dong Zhang[2]*

[1]School of Materials Science and Engineering, Shandong University, Jinan 250061, Shandong Province, China
[2]College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, Hunan Province, China

***Corresponding author:** Xu-dong Zhang, zxdcn3721@sina.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: By using HDI and TMP as the main raw materials, polyethylene glycol 400 (PEG400) is used as a non-ionic hydrophilic modifier, and sodium hydroxyethyl sulfonate is used as an ionic hydrophilic modifier to synthesize a dual hydrophilic modified polyurethane curing agent. Research revealed that introducing PEG400 for hydrophilic chain segments and sodium hydroxyethyl sulfonate for hydrophilic ionic groups in the polyurethane curing agent component leads to a uniform distribution of hydrophilic components, significantly enhancing compatibility with the aqueous polyol component, and results in excellent film performance. The synthesis process and film were characterized using Fourier transform infrared spectroscopy and high-resolution scanning electron microscopy in the study.

Keyword: Coating; Water polyurethane; Curing agent; Two-package; Double modification

Online publication: January 23, 2024

1. Introduction

Waterborne polyurethane with two components is not only environmentally friendly but also exhibits a low film-forming temperature, high film gloss, good wear resistance, water resistance, and solvent resistance. It currently stands as a research hotspot in waterborne coatings [1]. The two-component waterborne polyurethane consists of a polyisocyanate curing agent component and a water-based polyol component, which are uniformly mixed before coating. During film formation, the NCO in the curing agent reacts with the -OH in the polyol to form a cross-linked film.

In the realm of two-component waterborne polyurethane, hydrophilic-modified polyisocyanate finds widespread use, with two main categories: non-ionic modification and ionic modification. Non-ionic hydrophilic modification primarily employs polyether modification, introducing hydrophilic non-ionic ether segments into the molecular structure of polyisocyanate by reacting polyisocyanate with polyethylene glycol monomethyl ether. While this imparts hydrophilicity, it also introduces challenges related to water resistance and crystallinity tendencies [2-6]. Ionic modification includes cationic modification and anionic modification. Cationic modification introduces tertiary amine groups on the polyisocyanate molecule, obtained through acid

neutralization to form salts. However, cations have a promoting effect on the reactivity of NCO groups with active hydrogen, accelerating the reaction rate with water, leading to decreased system stability, multiple steps, and high cost [7,8]. Anionic modification introduces carboxyl or sulfonic acid groups on the polyisocyanate molecule, obtained through neutralization with a base, resulting in a curing agent with a pH of less than 7, which slows down the reaction rate of NCO with water, but additional neutralization steps are still required, adding complexity [9-11].

This study uses polyethylene glycol 400 (PEG400) and sodium hydroxyethyl sulfonate to perform non-ionic and ionic composite hydrophilic modification on hexamethylene diisocyanate-trimethylol propane (HDI-TMP) prepolymers. This approach combines the advantages of both hydrophilic modification methods, compensating for their respective shortcomings, and does not necessitate neutralization agents, thereby simplifying the process. The synthesized curing agent, studied in this research, exhibits a high NCO content, easy dispersal in water, and excellent compatibility with the water-based polyol component, resulting in outstanding coating performance.

2. Experiment

2.1. Materials

Hexamethylene diisocyanate (HDI) was purchased from Wanhua Chemical Group Co., Ltd. (China). Polyethylene glycol 400 (PEG400) was obtained from Sinopharm Medicine Holding Group Co, Ltd. (China). Trimethylolpropane (TMP) was purchased from LookChem Chemical Group Co., Ltd. (China). Sodium 2-hydroxyethyl sulfonate (SHS; purity > 98%) was provided by Shanghai Jingchun Co., Ltd. (China). Polyurethane polyol components (hydroxyl value = 26.63 mgKOH/g) were self-made.

2.2. Synthesis of curing agent

The appropriate amount of HDI was added into a four-neck flask, with a condensing tube, thermometer, and stirring paddle. The TMP solution was then dripped into the flask within 1 h at 40℃ under stirring. After that, the mixture was kept at 80℃ for 4.5 h, during this process, PEG400 was added to the mixture, followed by keeping the mixture at 73℃ for 1.5 h with the addition of sodium hydroxyethyl sulfonate. Finally, the double hydrophilic modified polyurethane curing agent was obtained. The reaction scheme is illustrated in **Figure 1**.

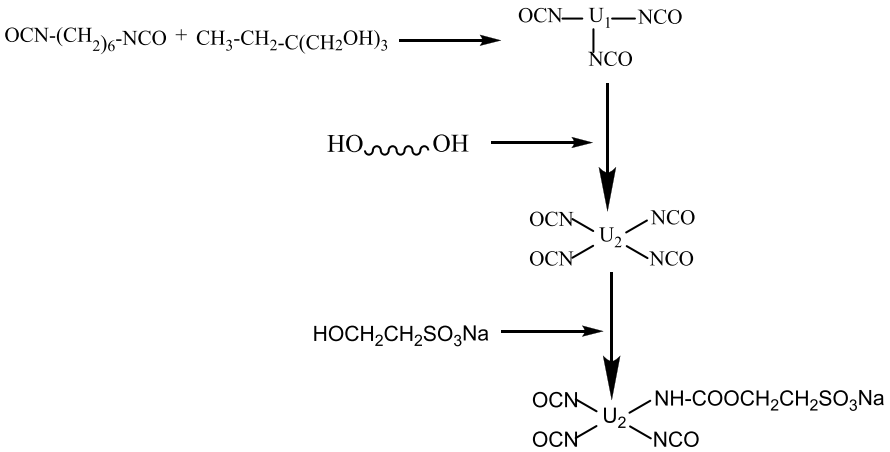


Figure 1. The synthesis process of curing agent.

2.3. Preparation of the film

The curing agent and polyurethane polyol components were mixed well with the ratio of , and then the emulsion was coated on the steel substrate. The film was stored at room temperature for 2 h and then dried at 50°C for 24 h.

2.4. Characterization

The molecular structure and chemical composition of the curing agent were characterized using the Fourier transform infrared spectrometer (WQF-410). The hardness of the films was measured using a pencil hardness tester, the average value of three times was taken as the final pencil hardness. Gloss measurement was performed with a QZX-60A gloss meter (Tianjin material test factory). Ethanol resistance was tested using the drop method. Drops of 50% (volume fraction) ethanol solution were dropped on the paint film and covered with a glass surface dish or a plastic cover to prevent the liquid from evaporating too quickly. The paint was cleaned with water or a clean cloth after being placed at room temperature for 1 h, and the degree of coloring of the liquid on the paint was observed. Water absorption was evaluated in terms of the difference between the wet mass (W_1) and dry mass (W_2) of the coating, and was calculated using the following equation:

3. Result and discussion

3.1. Fourier transform infrared spectroscopy (FTIR) analysis

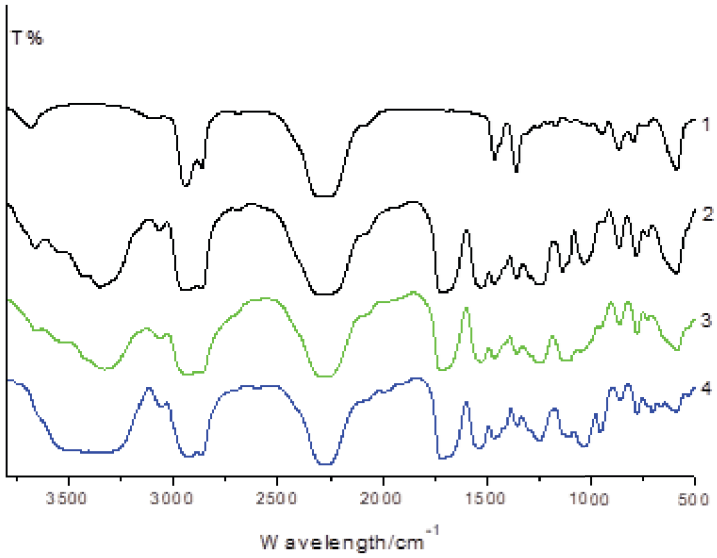


Figure 2. FTIR spectra of the progress of the polyurethane curing agent synthesis. **(1)** The material of HDI; **(2)** The product with TMP reaction for 3 hours; **(3)** The polymer after reacting with PEG400 for 1.5 h; **(4)** The curing agent

Figure 2 shows the synthesis progress of the polyurethane curing agent. From the infrared absorption spectrum of HDI, the peak at 2,304 cm^{-1} is ascribed to the –NCO group. After the reaction between HDI and TMP, the N–H stretching vibration peak is observed at about 3,331 cm^{-1}, which indicates that the TMP has successfully reacted with HDI. During the synthesis progress, the peak of the N–H group widens gradually, which indicates that TMP and PEG400 have been introduced into the macromolecules of the curing agent. Besides, the peaks at 1,203 cm^{-1} and 1,052 cm^{-1} represent the S=O asymmetric and symmetrical stretching vibration, which indicates that the sodium hydroxyethyl sulfonate has reacted with the precursor.

3.2. The effect of reaction time and reaction temperature of HDI and TMP

According to the reaction equation between HDI and TMP, as the reaction progresses, the proportion of –NCO groups gradually decreases. The percent of –NCO groups decrease is displayed in **Figure 3**, which is detected by the toluene-di-n-butylamine method. From **Figure 3**, in the case of the same ratio of HDI and TMP, the percent of –NCO groups in the structure decreases with the extension of reaction time. When the reaction was conducted at 80℃ for 3 h, the NCO content was close to the theoretical content and then remained unchanged. When the reaction is under 70℃, the theoretical content can be reached at 70℃ after a 5-hour reaction. At 90℃, the NCO content was close to the theoretical content for the initial 3 h but subsequently decreased. This reduction is mainly attributed to the elevated temperature causing carbamate group generation, leading to a side reaction with the NCO group and consequently consuming the NCO group. It is evident from this experiment that the optimal reaction conditions are set at 80℃ for 3 h.

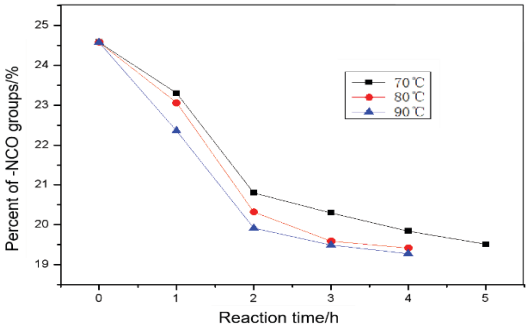


Figure 3. The relationship of NCO content with temperature and reaction time

3.3. The ratio of HDI and TMP analysis

The theoretical molar ratio of the synthesized HDI-TMP prepolymer is 3:1 (). However, due to the occurrence of side reactions, it was necessary to control in practical application. The effects of ratios on the coating properties were studied as shown in **Table 1**. It indicated that both gloss and hardness were enhanced with the increasing of . This is because the increasing of made the TMP reaction completely and HDI-TMP prepolymer molecular structure tended to idealize. However, too high ratio leaded to too much free NCO groups, which were directly modified by polyethylene glycol and sulfonate, resulting in not only decreased cross-linking density, water resistance, and alcohol resistance but also increased water absorption. Therefore, the appropriate for this experiment was about 3.0.

Table 1. The effect of ratio on the film properties

	2.1	**2.5**	**3.0**	**3.5**	**4.0**
Gloss (60°)	103.8	105.2	112.7	115.4	120.9
Hardness	2H	3H	3H	3H	3H
Flexibility (mm)	1	1	1	1	1
Adhesion	0	0	0	0	0
Impact resistance (kg/cm)	45	45	45	45	45
Water absorption (%)	6.8	9.3	14.8	20.1	25.9
Resistance to ethanol	No change	Micro-whitened	Micro-whitened	Whitened	Whitened
Mixed state with polyol dispersion	Good	Good	Good	Good	Good

3.4. PEG400 content analysis

Table 2 shows the properties of the film with the additions of PEG400 in the curing agent under the condition that and sodium hydroxyethyl sulfonate dosage remains unchanged. It was easily found that the gloss, water absorption of film, and compatibility with polyurethane polyol dispersion were further enhanced according to the molecular weight and hydrophilicity of the PEG400. In addition, the introduction of polyethylene glycol into the molecular structure of the curing agent as a soft segment had a great impact on the flexibility of the coating film. All things considered, the optimal dosage of PEG400 in this experiment is 15.8%.

Table 2. The effect of PEG400 content on the curing agent and film properties

PEG content (w%)	12.1	14.2	15.8	17.0	18.0
Outside	Pale yellow & translucent	Pale yellow & translucent	Pale yellow & translucent	Pale yellow & translucent	Pale yellow & translucent
Viscosity of the curing agent(s)	52	56	59	64	67
Gloss (60°)	107.8	110.7	117.3	120.8	126.6
Hardness	3H	3H	3H	3H	3H
Flexibility (mm)	2	1	1	1	1
Adhesion	0	0	0	0	0
Impact resistance (kg/cm)	50	50	50	50	50
Water absorption (%)	6.7	9.8	13.1	19.8	22.4
Mixed state with polyol dispersion	Floccule	Floccule	Good	Good	Good

3.5. SHS content analysis

SHS was introduced into the curing agent as the anionic hydrophilic modifier because of the strong hydrophilicity of sulfonate groups which led to an increase in water absorption of the coating films as shown in **Table 3**. Additionally, the surface of the films was smooth and compact due to the good compatibility of the curing agent and polyethylene glycol, which could be responsible for the improvement of the gloss of the film with the increase of sodium hydroxyethyl sulfonate content. In order not only to ensure good compatibility with polyol components but also to make the water absorption relatively low, the appropriate amount of SHS is about 3.17%.

Table 3. The effect of sodium hydroxyethyl sulfonate content on the film properties

Content (%)	3.06	3.17	3.40	3.60	3.78
Gloss (60°)	106.3	110.5	114.2	126.1	128.9
Hardness	3H	3H	3H	3H	3H
Flexibility (mm)	1	1	1	1	1
Adhesion	0	0	0	0	0
Impact resistance (kg/cm)	50	50	50	50	50
Water absorption (%)	5.4	9.6	14.4	16.8	19.3
Mixed state with polyol dispersion	Floccule	Good	Good	Good	Good

3.6. The effect of

The influence of polyurethane curing agents and self-made polyurethane polyol components on film

properties was studied by changing different ratios, and the results are shown in **Table 4**. With the increase of the ratio, the hardness of the film increases, the water resistance and solvent resistance improve, and the gloss increases initially and then decreases. The main reason was that with the increase of the curing agent dosage, the content of the carbamate group and urea group generated by the reaction of the curing agent and polyol component changed, which led to the increase in film hardness. Simultaneously, the amount of curing agent is directly proportional to the crosslinking density of the film, thereby the film becomes smoother and brighter. Nevertheless, more urea groups and CO_2 are generated, and as the compatibility of the urea group and carbamate group is poor, this leads to reduced gloss.

Table 4. The effect of ratio on the film properties

	1.2	**1.4**	**1.6**	**1.8**	**2.0**
Surface drying time (min)	32	37	41	47	50
Gloss (60°)	107.8	109.5	114.3	118.3	110.4
Hardness	3H	3H	3H	4H	4H
Flexibility (mm)	1	1	1	1	1
Adhesion	0	0	0	0	0
Impact resistance (kg/cm)	50	50	50	50	50
Water absorption (%)	74.6	67.1	23.5	16.4	9.2
Resistance to ethanol	Serious whitened	Whitened	Whitened	Micro-whitened	Micro-whitened

Figure 4 shows the scanning electron microscopy of uncured polyurethane polyols and coating films with different proportions. It can be seen that the surface of the uncured polyol coating is not smooth and has large cracks. With the increase of the ratio, the surface of the coating becomes more regular and smooth, the cracks become smaller, and the densification becomes better, which is also consistent with the changes in hardness, water resistance, and other properties tested in **Table 4**.

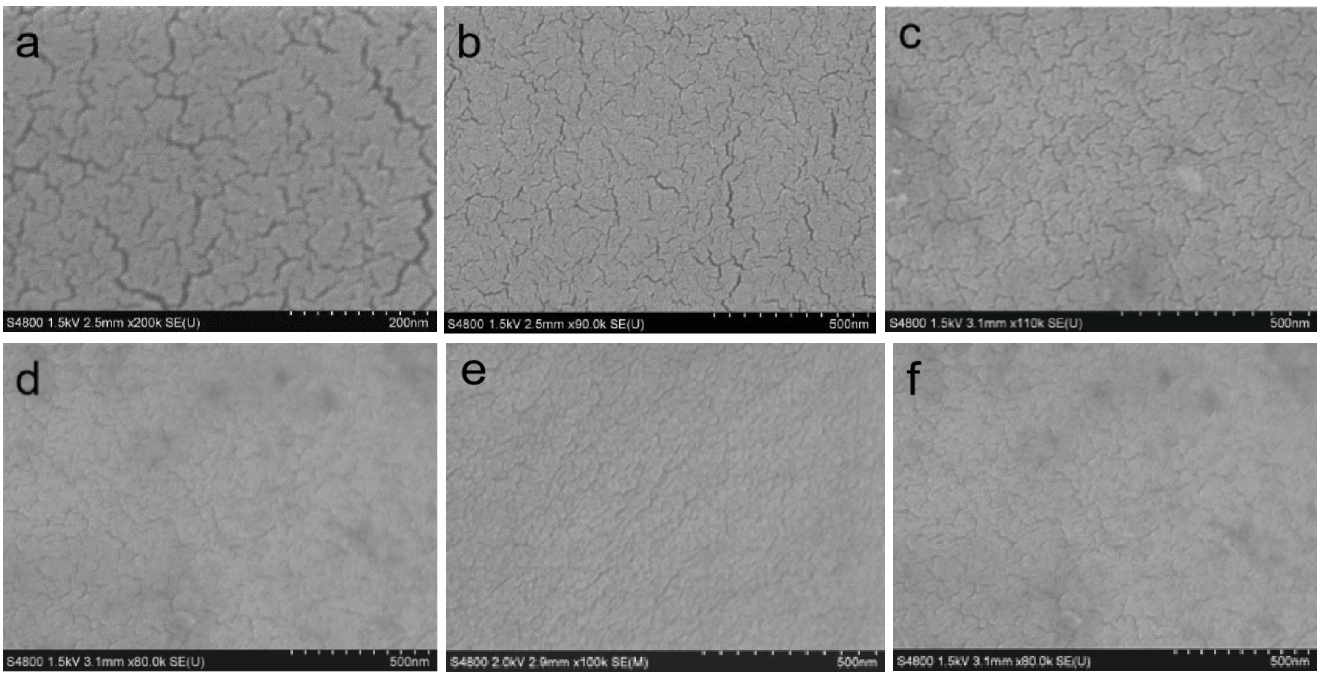


Figure 4. The SEM image of films. **(a)** polyurethane polyols; **(b)** ; **(c)** ; **(d)** ; **(e)** ; **(f)**

4. Conclusion

A series of double hydrophilic modified waterborne polyurethane curing agents were prepared with HDI/TMP/PEG400/SHS, which had better compatibility with polyol components. The WPU films had good flexibility, gloss, and low water absorption when the ratio of HDI/TMP was 3.0, with an optimal dosage of PEG400 and SHS of 15.8% and 3.17%, respectively.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Melchiors M, Sonntag M, Kobusch C, et al., 2000, Recent Developments in Aqueous Two-Component Polyurethane (2K-PUR) Coatings. Progress in Organic Coatings, 40(1–4): 99–109. https://doi.org/10.1016/S0300-9440(00)00123-5

[2] Li K, Shen Y, Fei G, et al., 2014, The Effect of PETA/PETTA Composite System on the Performance of UV Curable Waterborne Polyurethane Acrylate. Journal of Applied Polymer Science, 132(2): 41262. https://doi.org/10.1002/app.41262

[3] Wu J, Yan H, Wang J, et al., 2013, Flame Retardant Polyurethane Elastomer Nanocomposite Applied to Coal Mines as Air-Leak Sealant. Journal of Applied Polymer Science, 129(6): 3390–3395. https://doi.org/10.1002/app.38946

[4] Liao K-H, Park YT, Abdala A, et al., 2013, Aqueous Reduced Graphene/Thermoplastic Polyurethane Nanocomposites. Polymer, 54(17): 4555–4559. https://doi.org/10.1016/j.polymer.2013.06.032

[5] Liu J, Liu H-F, Deng Li, et al., 2013, Improving Aging Resistance and Mechanical Properties of Waterborne Polyurethanes Modified by Lignin Amines. Journal of Applied Polymer Science, 130(3): 1736–1742. https://doi.org/10.1002/app.39267

[6] Schwindt J, Reiff H, Kubitza W, 1995, Coating Compositions, A Process for Their Production and Their Use for Coating Water-Resistance Substrates. US5459197A, United States, viewed April,5nd,2023, https://patents.google.com/patent/US5459197A/en

[7] Kim BK, Lee JC, 1996, Waterborne Polyurethanes and Their Properties. Journal of Polymer Science Part A: Polymer Chemistry, 34(6): 1095–1104. https://doi.org/10.1002/(SICI)1099-0518(19960430)34:6<1095::AID-POLA19>3.0.CO;2-2

[8] Noble K-L, 1997, Waterborne Polyurethanes. Progress in Organic Coatings, 32(1–4): 131–136. https://doi.org/10.1016/S0300-9440(97)00071-4

[9] Mosbach J, Laas H-J, Jubitza W, 1992, Polyisocyanate Mixtures, Processes for Their Production and Their Use as Binders for Coating Compositions or as Reactants for Compounds Reactive to Isocyanate Groups or Carboxyl Groups. US5098983A, United States, viewed August,2nd,2023, https://patents.google.com/patent/US5098983A/en

[10] Haberle K, 1996, Water-Emulsifiable Polyisocyanates. US5583176A, United States, viewed Jan,13nd,2023, https://patents.google.com/patent/US5583176A/en

[11] Wang L, Lai XJ, Ma SY, 2013, Preparation of UV Curable Waterborne Polyurethane and its Application in Coatings. Advanced Materials Research, 821–822: 925–928. https://doi.org/10.4028/www.scientific.net/AMR.821-822.925

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Design of Cold-Junction Compensation and Disconnection Detection Circuits of Various Thermocouples and Implementation of Multi-Channel Interfaces Using Them — A Secondary Publication

Hyeong-Woo Cha*

Department of Electronic Engineering, Cheongju University, Korea

***Corresponding author:** Hyeong-Woo Cha, hwcha@cju.ac.kr

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Cold-junction compensation (CJC) and disconnection detection circuit design of various thermocouples (TC) and multi-channel TC interface circuits were designed. The CJC and disconnection detection circuit consists of a CJC semiconductor device, an instrumentation amplifier (IA), two resistors, and a diode for disconnection detection. Based on the basic circuit, a multi-channel interface circuit was also implemented. The CJC was implemented using compensation semiconductor and IA, and disconnection detection was detected by using two resistors and a diode so that IA input voltage became -0.42 V. As a result of the experiment using R-type TC, the error of the designed circuit was reduced from 0.14 mV to 3 μV after CJC in the temperature range of 0°C to 1400°C. In addition, it was confirmed that the output voltage of IA was saturated from 88 mV to -14.2 V when TC was disconnected from normal. The output voltage of the designed circuit was 0 V to 10 V in the temperature range of 0°C to 1400°C. The results of the 4-channel interface experiment using R-type TC were almost identical to the CJC and disconnection detection results for each channel. The implemented multi-channel interface has a feature that can be applied equally to E, J, K, T, R, and S-type TCs by changing the terminals of CJC semiconductor devices and adjusting the IA gain.

Keywords: R-type thermocouple (TC); Cold-junction compensation (CJC); TC disconnection detection; Multi-channel interface circuit; Sensor interface

Online publication: January 29, 2024

1. Introduction

High-temperature, high-precision heat treatment devices are rapidly increasing in the industries of semiconductors, new materials, and MEMS (Micro Electro Mechanical Systems). In these heat treatments, it is important to have a temperature control device that maintains a constant temperature in the chemical processing

area (e.g., 300 mm wafer oxidation and deposition process) and verifies the presence of abnormalities (breaks, short circuits).

Although the localization of various semiconductor manufacturing equipment is actively taking place in Korea, in the case of electric furnaces, high-temperature heaters and temperature controllers are imported and the entire electric furnace is localized. For the temperature control of an electric furnace for 300 mm wafers, it is necessary to control the temperature in a quartz tube with a diameter of 400 mm and a height of 1.5 m to ensure that the set temperature is constant in the range of 0°C to 1400°C. In this case, the tube is divided into four zones, and thermocouple (TC) temperature sensors are used for precise temperature control.

Figure 1 shows the voltage characteristics generated by various TCs as the temperature changes. From this figure, it can be seen that TCs of types R, S, and B should be used for measuring temperatures ranging from 0°C to 1400°C. Since all TCs are temperature sensors that generate voltages at arbitrary temperatures with 0°C as the reference, the electromotive force must be corrected for 0°C, which is the cold junction for room temperature (25°C) [1].

A primitive method of cold-junction compensation is to use ice water to measure electromotive force by placing TC1 outside (the environment to be measured) and TC2 inside the ice, as shown in **Figure 2**. In this case, it is difficult to apply in practice because the measurement system is complex, has a large appearance, and is particularly difficult to supply ice. These problems have recently been solved using zero-calibration semiconductor chips [1-3].

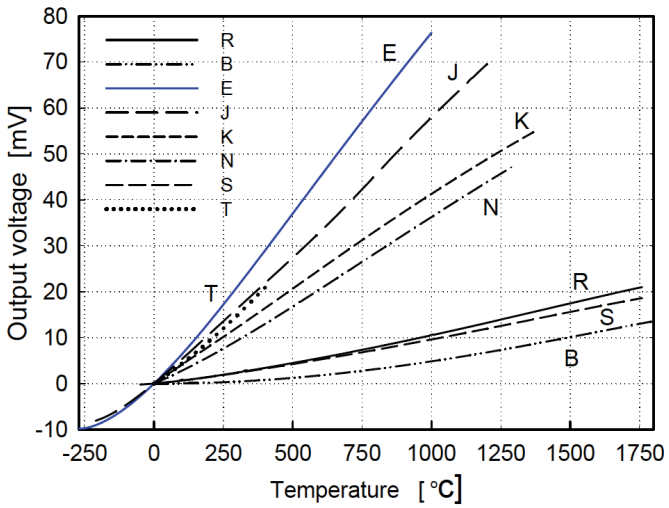


Figure 1. Voltage characteristics generated according to temperature for each type of TC

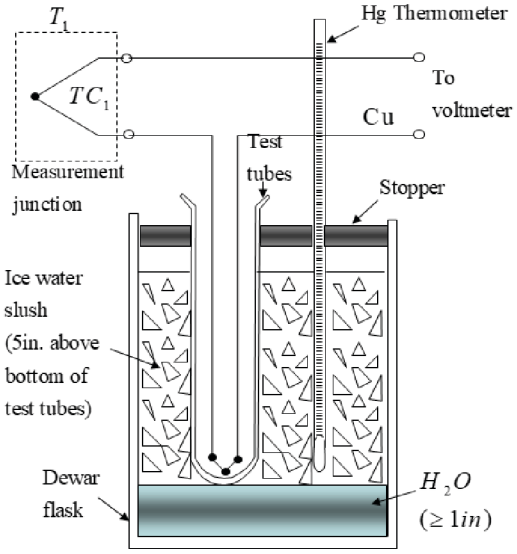


Figure 2. Physical coolness correction principle of TC

In addition, since the TC is open due to degradation when used at high temperature for a long time, it is essential to have a function to detect the breakage of the TC in order to quickly maintain the heat treatment device. In general, electric furnaces used in semiconductor processes use four or more multichannel TCs, so it is possible to use a high-performance microprocessor to measure the temperature, check for breaks, and linearize the TCs by program coding.

Prior to this study, we presented a cold-junction compensation (CJC) and disconnection detection circuit of R-type TC sensor for localization of temperature controller of electric furnace for 300 mm or larger wafers [4]. In this paper, the applicability of various TCs was not confirmed, and it was suggested that multi-

channel application was possible for R-type TCs. Since then, Microchip has developed a dedicated chip that can use various TCs (e.g., K, J, T, N, S, E, B, R-type) and commercialized it as a temperature-to-digital signal conversion module [6-8]. In this case, it was possible to process signals from multiple TCs through I^2C communication. However, in this case, depending on the number of TCs used, a conversion module is required, and additional circuits are required to detect disconnection.

In this study, we investigate whether CJC and disconnection detection circuit can be applied not only to R-type TC sensors but also to E, J, K, T, and S-type TCs, and design a complementary design. Based on the designed circuit, a cold compensation and open wire detection interface that can be used up to multi-channel (using n TCs) is designed and its performance is analyzed.

2. Circuitry and how it works

2.1. Behavioral characteristics of TCs

As shown in **Figure 3**, a TC is a device that generates electromotive force by the seedback effect when two dissimilar materials are joined and then subjected to a temperature difference, and the electromotive force is proportional to the temperature difference between the cold junction of 0°C and any temperature. The electromotive force of TC is given by the following formula [3].

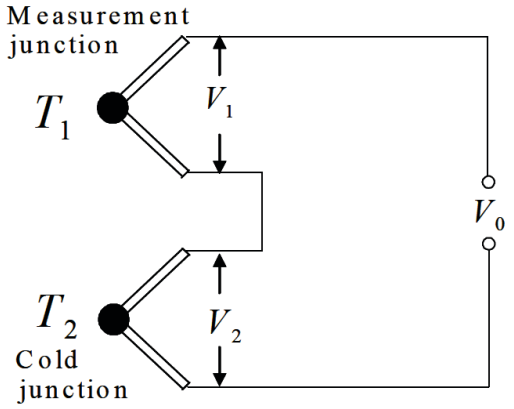


Figure 3. Principle diagram of voltage generation in TC

$$V_O = V_1 + V_2 = \alpha(T_1 - T_2) + \gamma(T_1^2 - T_2^2) \quad (1)$$

Here, α and ɣ are constants given by the junction material and T_1 and T_2 are the respective temperatures for V_1 and V_2. From **Figure 3** and Equation (1), it can be seen that if we make $T_2 = 0$°C, $V_2 = 0$V, and Equation (1) is given by $V_0 = V_1 = \alpha T_1 + \gamma$. In addition, at room temperature (20°C), $T_1 = T_2$ and $V_0 = 0$, so the TC sensor cannot measure the room temperature. Therefore, the voltage compensation of the TC for the room temperature, that is, the cold-junction compensation (CJC) must be performed so that V_2 is 0V.

TCs differ in their measurable temperature range, magnitude of electromotive force, and linear characteristics depending on the material being junctioned. In the case of electric furnaces, R-type TCs are commonly used because they are always in a standby state and the process time is repeated. The R-type TC generates electromotive force from 0 mV to 16.04 mV from 0°C to 1400°C and has a temperature sensitivity of 11.457 μV/°C.

2.2. Working principle of cold-junction compensation (CJC) semiconductor devices

Figure 4 shows the pin designation (a) and internal block diagram (b) of the TC CJC semiconductor device

(LT1025). The chip has four outputs of 60.9 μV/°C (E-type TC), 51.7 μV/°C (J-type TC), 40.6 μV/°C (K and T-type TC), and 6 μV/°C (R and S-type TC) [5]. The internal block consists of a temperature sensor with a sensitivity of 10 mV/°C, a bow correction voltage, an amplifier (buffer), and a TC-specific temperature sensitivity (slope) adjustment resistor.

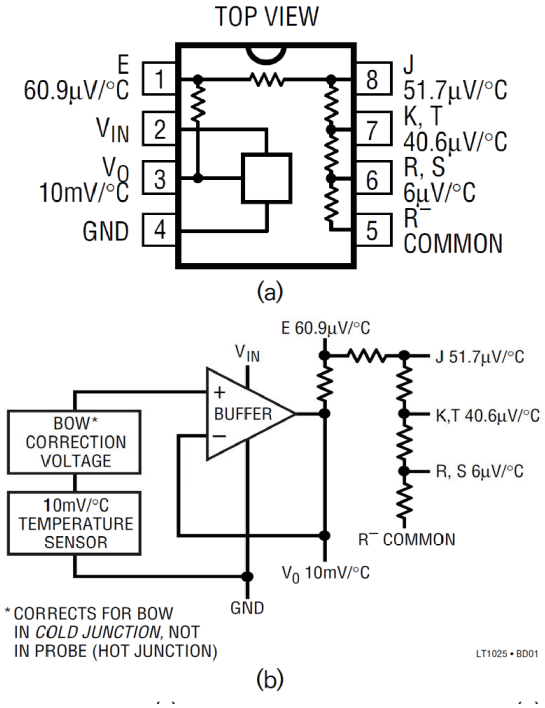


Figure 4. (a) Pin name and (b) internal block diagram of TC's CJC semiconductor chip (LT1025)

2.3. Conventional K-type TC cold compensation interface circuit

The cold compensation circuit of a conventional K-type TC is shown in **Figure 5** [5]. The circuit consists of a cold compensation device LT1025, a K-type TC, and a voltage amplifier [4]. This circuit has a sensitivity of 10 mV/°C for the output voltage depending on the temperature, but there is a problem that the output voltage becomes 0V when the TC is disconnected, so it is not possible to distinguish whether the output of the TC is actually at 0°C or disconnected. In particular, no compensation circuit for R-type TCs has been announced, and no disconnection detection circuit has been announced.

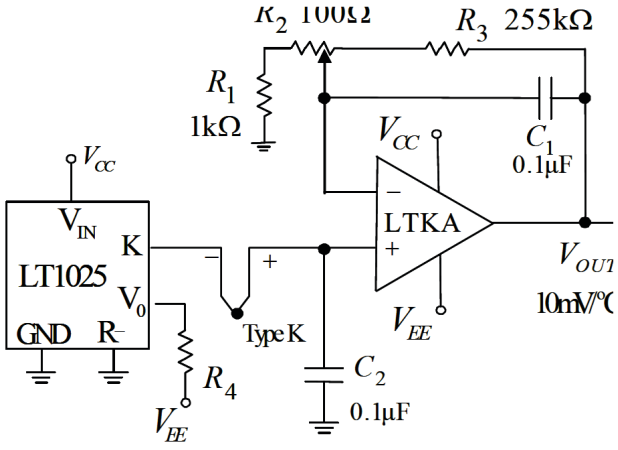


Figure 5. Schematic of conventional K-Type TC interface [5]

2.4. Design of cold-junction compensation and disconnection detection circuits of various TCs and multi-channel application circuits

2.4.1. Design of cold-junction compensation and disconnection detection circuit for various TCs

A circuit to detect cold-junction compensation and TC disconnection for various TCs was designed, and the designed circuit is shown in **Figure 6**. The circuit consists of 1) a cold-junction compensation device LT1025, an R-type TC, a gain adjustment resistor RG, and an instrument amplifier (IA) for cold-junction compensation, and 2) a circuit consisting of R_1, R_2, D_1, and IA for disconnection detection.

The principle of cold-junction compensation is as follows. In other words, if only the first term in the TC electromotive force characteristic in Equation (1) is considered, the electromotive force of an R-type TC sensor under room temperature conditions can be represented by Equation (2).

$$\begin{aligned} V_{TC}(R) &= \alpha(T_P - T_{ROOM}) + \alpha T_{CJC} \\ &= \alpha T_P - \alpha T_{ROOM} + \alpha T_{CJC} \end{aligned} \qquad (2)$$

Where T_P is the temperature of any space where the TC sensor measures the temperature (in the case of an electric furnace, the temperature in the quartz tube), T_{ROOM} is the room temperature, and T_{CJC} is the cold-junction compensation temperature.

Generally, the arbitrary space has a temperature corresponding to room temperature, so the electromotive force of the TC sensor is 0V because $TP = T_{ROOM}$ is established at room temperature. However, since the TC generates electromotive force for any temperature difference relative to 0°C, it is necessary to add the electromotive force αT_{CJC} corresponding to room temperature, that is, $-\alpha T_{ROOM} + \alpha T_{CJC} = 0$. Therefore, if R_2 is very large in the circuit shown in **Figure 6**, this resistor can be considered open, so $V_{IN} = \alpha T_P$, and the electromotive force equal to the value in the characteristic table of the R-type TC is obtained. This electromotive force is the input to the instrumentation amplifier (AD620), so the final output voltage V_{OIA} of the cold-junction compensation circuit is given as follows [10].

$$V_{OIA} = \left(\frac{49.4\text{k}}{R_G} + 1\right)\alpha T_P \qquad (3)$$

Since the electromotive force is 0 mV–16.04 mV for the R-type TC in the temperature range of 0°C–1400°C, using the above equation, $R_G = 79.4\ \Omega$ (gain is 623.44) can be used to obtain an output voltage of 0–10 V, and $R_G = 158.84\ \Omega$ (gain is 311.72) can be used to obtain an output voltage of 0–5 V.

In the circuit diagram of TC cold compensation and disconnection detection shown in **Figure 6**, R (pin 6) of the LT1025 chip is used for R-type TC, but E (pin 1), J (pin 8), K and T (pin 7), and S (pin 6) terminals of the LT1025 chip shown in **Figure 4** can be used to perform cold-junction compensation and disconnection detection for each TC at the same time by using E, J, K, T, and S-type TC. In other words, if other TCs are used as shown in **Figure 6**, the value of R_G, which adjusts the gain of the measurement amplifier in the dotted circuit, must be changed, and the terminals of the LT1025 chip can be used for that TC.

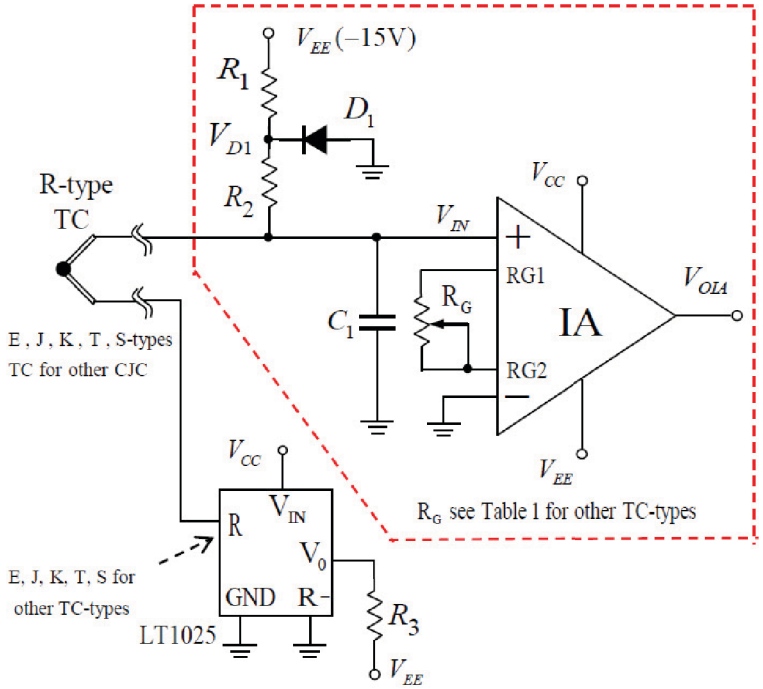


Figure 6. Cold-junction compensation and disconnection detection circuit diagram of various TCs [4]

Since the TCs have different sensitivities to temperature and different electromotive forces at room temperature (25°C), the gain of the measurement amplifier must be adjusted separately. For various TCs, the measuring temperature range, electromotive force range, gain of the measurement amplifier, and R_G value for each TC to obtain 0–10 V output of the measurement amplifier are shown in **Table 1**.

Table 1. Measuring temperature range and output voltage of TCs, gain of the instrumentation amplifier (IA), and R_G of IA for each TC to obtain the output voltage of 0–10 V of IA

TC type	Temperature range (°C) [6]	Output voltage of TC (mV) [6]	Gain of IA	R_G (Ω)
E	0–1,000	0–76.37	130.09	382.68
J	0–1,200	0–69.55	143.78	345.98
K	0–1,350	0–54.82	182.41	272.31
T	0–400	0–20.87	479.16	103.31
S	0–1,760	0–18.61	537.34	92.11

2.4.2. Principle of TC disconnection detection circuit

If the TC is disconnected, the circuit in **Figure 6** can be modified as shown in **Figure 7** [4]. In this circuit, D_1 is forward through -15 V (V_{EE}), R_1, D_1, and ground, resulting in $V_{D1} \cong$ -0.6 V (forward voltage of the diode). When TC is disconnected, the voltage accumulated on C1, $V_{IN} = \alpha T_P$, discharges to -0.6 V through R_1 and R_2, so V_{IN} can be represented as follows.

$$V_{IN} = a T_P - \frac{(a T_P + V_{D1})}{C_1 (R_1 + R_2)} t \qquad (4)$$

From this expression, it can be seen that $V_{IN} \simeq -V_D$ when time t becomes equal to the time constant $C_1(R_1+R_2)$. For R-type TCs, V_{IN} will have a value close to -0.6 V because the maximum value of αT_P is about 16 mV. This voltage is converted to a negative saturation voltage by the measurement amplifier IA. Therefore, since the temperature change value of the TC is converted to a positive voltage, it can be judged that the TC is disconnected when the measurement amplifier output voltage is converted from a positive voltage to a negative saturation voltage. In other words, if the output V_{OIA} of the measurement amplifier is a negative saturation voltage, the TC is disconnected. Therefore, the circuit shown in **Figure 7** can be used to detect the breakdown of various TCs, namely E, J, K, T, and S-type TCs. However, the value of the gain adjustment resistor R_G of the measuring amplifier should be adjusted as shown in **Table 1** according to the TC.

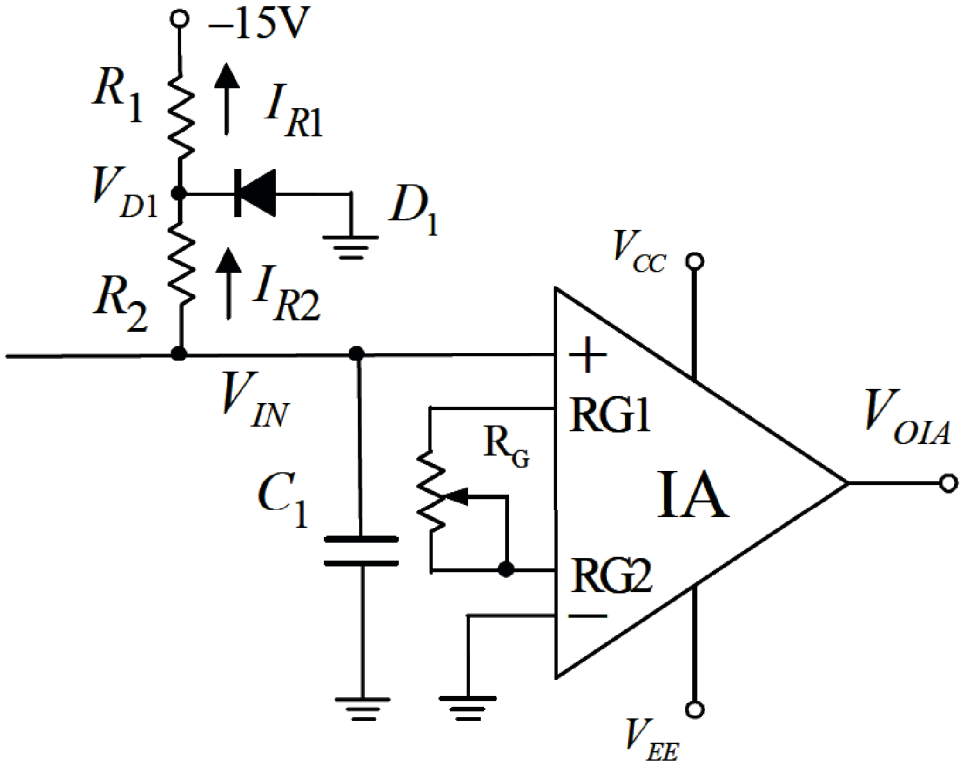


Figure 7. Circuit diagram explaining the TC open detection principle

2.4.3. Interface circuit design for multi-channel, various TC cold-junction compensation and disconnection detection circuit

Based on the cold-junction compensation (CJC) of various TCs and the TC disconnection detection circuit shown in **Figure 6**, multi-channel TC cold-junction compensation and disconnection detection can be designed for various TCs, and the interface circuit applicable to multi-channel R-type TCs or E, J, K, T, S-type TCs is shown in **Figure 8**. The circuit consists of n TCs, one cold-junction compensation semiconductor, n break detection circuits and instrumentation amplifiers, a multiplexer (MUX) to select one of the multichannel TCs, and an instrumentation amplifier to amplify the final input signal to an analogue-to-digital (A/D) converter.

In **Figure 8**, the R terminal of the LT1025 is connected as a TC, so TC1 to TCn can be used as R-type TCs. When used as a multichannel interface to E, J, K, T, and S-type TCs, the E, J, K, T, and S terminals of the LT1025 device (**Figure 4**) can be used and the corresponding TCs can be used. Moreover, since the

electromotive force is different depending on the TC, the value of R_3 inside the dotted line in **Figure 8** can be selected as the resistor value shown in **Table 1**.

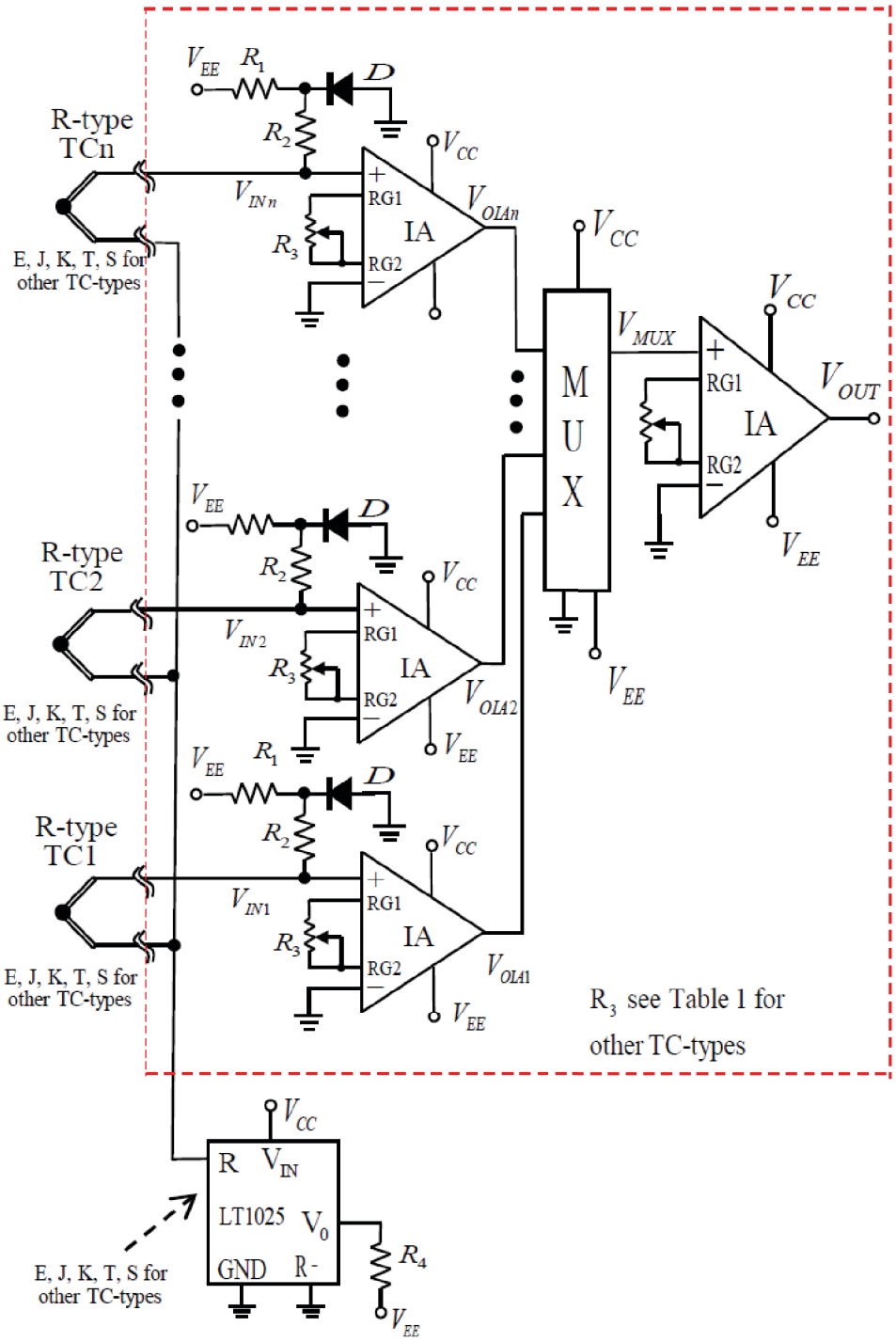


Figure 8. Interface circuit applicable to multi-channel R-type TC or E, J, K, T, S-type TC based on Figure 6

Using the multichannel (n) cold-junction compensation and disconnection detection circuit shown in **Figure 8**, which is implemented with a measurement system consisting of a 16-bit or higher A/D converter and a microprocessor, it is possible to measure the temperature at n different locations and detect whether the TC used at each location is disconnected.

3. Results

The designed **Figures 6** and **8** were experimented using individual devices. The devices used are D_1 = IN4937, IA = AD 620, MUX = ADG406, CJC device = LT1025, R_1 = 10 MΩ, R_2 = 50 MΩ, R_G = 79.4 Ω, C_1 = 1 μF. The supply voltage was set to V_{cc} = 15 V and V_{EE} = -15 V. In addition, among the E, J, K, T, R, and S-type TCs in **Figures 6** and **8**, the R-type TC of Sentech, which is widely used in electric furnaces and is semiconductor processing equipment, was used [2,9]. In the circuit in **Figure 8**, the circuit for 4 channels was verified through experiments with n = 4.

Figure 9 shows the difference between the input voltage V_{IN} of the measurement amplifier and the electromotive force output given in the characteristic table of the R-type TC for the case of cold-junction compensation in the circuit shown in **Figure 6**. The temperature-dependent electromotive force of the R-type TC was used as an input signal using YOKOGAWA's TC Calibrator CA320 [10-12]. In the graph in **Figure 9**, the dotted line shows the difference between the V_{IN} and TC electromotive force before cold-junction compensation, and the solid line shows the difference between the V_{IN} and TC electromotive force after cold-junction compensation. From this result, it can be seen that the difference between the TC's electromotive force and the input voltage V_{IN} is 3 μV in the temperature range of 0°C–1400°C through cold-junction compensation. In other words, the error after zeroing the TC is reduced from 0.14 mV to 3 μV, which is a reduction of about 97.8%.

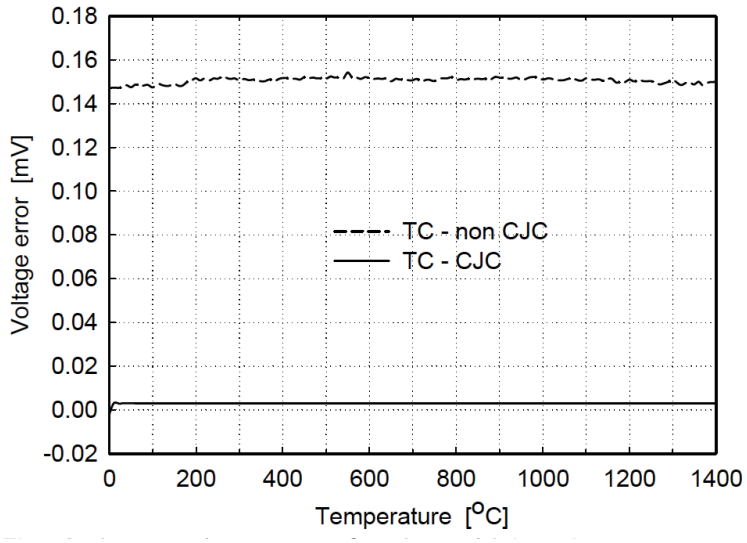


Figure 9. Input voltage error for the cold-junction compensation result of the circuit shown in Figure 6

Figure 10 shows the output voltage of the instrumentation amplifier (IA) with and without cold-junction compensation in the circuit shown in **Figure 6**. Here, R_G = 79.4 Ω was set to increase the gain by 632.44 times to output a voltage of 0 V to 10 V for temperatures from 0°C to 1400°C. In the graph in **Figure 10**, the solid line is the instrumentation amplifier output voltage before cold-junction compensation and the dashed line is the result of cold-junction compensation. From this result, it can be seen that the output voltage for the temperature range of 0°C–1400°C can be accurately obtained from 0–10 V by cold-junction compensation.

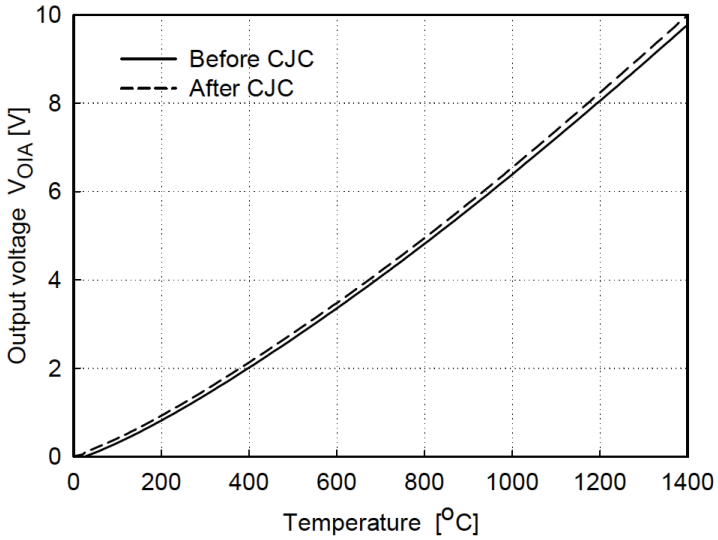


Figure 10. Output voltage of the result of cold-junction compensation for the circuit shown in Figure 6

In the circuit shown in **Figure 8**, the input voltage V_{IN} and output voltage V_{OIA1} to V_{OIA4} of the instrumentation amplifier (IA) are shown in **Tables 2** and **3**, respectively, when four R-type TCs are used in the 4-channel interface circuit and one is disconnected. From **Table 2**, it can be seen that the input V_{IN} is about -421 mV when each TC is disconnected, and the V_{IN} is about 0.14 mV when the TC is not disconnected (electromotive force of the TC at room temperature 25°C).

Table 2. Input voltage status of IA when TC is disconnected in the circuit of **Figure 8**

	TC1	**TC2**	**TC3**	**TC4**
Vin1 [mV]	-421.0	0.141	0.141	0.141
Vin2 [mV]	0.141	-420.5	0.141	0.141
Vin3 [mV]	0.141	0.141	-420.6	0.141
Vin4 [mV]	0.142	0.141	0.141	-420.5

From **Table 3**, it can be seen that the output voltage V_{OIA} is about -14.22 V when each TC is disconnected and about 87 mV (= 0.14 mV × 632.44) when it is not disconnected, as a result of amplifying the input signal of the TC by about 632.44 times with an instrumentation amplifier. Therefore, it is possible to measure V_{OIA} using an A/D converter and program the TC to be disconnected if a negative voltage appears.

Table 3. Output voltage status of IA when TC is disconnected in the circuit of **Figure 8**

	TC1	**TC2**	**TC3**	**TC4**
V_{OIA1} [V]	-14.215	0.087	0.088	0.087
V_{OIA2} [V]	0.087	-14.215	0.087	0.088
V_{OIA3} [V]	0.088	0.088	-14.215	0.088
V_{OIA4} [V]	0.087	0.087	0.087	-14.215

The actual photograph of the experimental environment is shown in **Figure 11**, where (a) the R-type TC of Sentech, (b) the experimental board of **Figure 8**, and (c) the TC Calibrator CA 320 of YOKOGAWA [2,11,12].

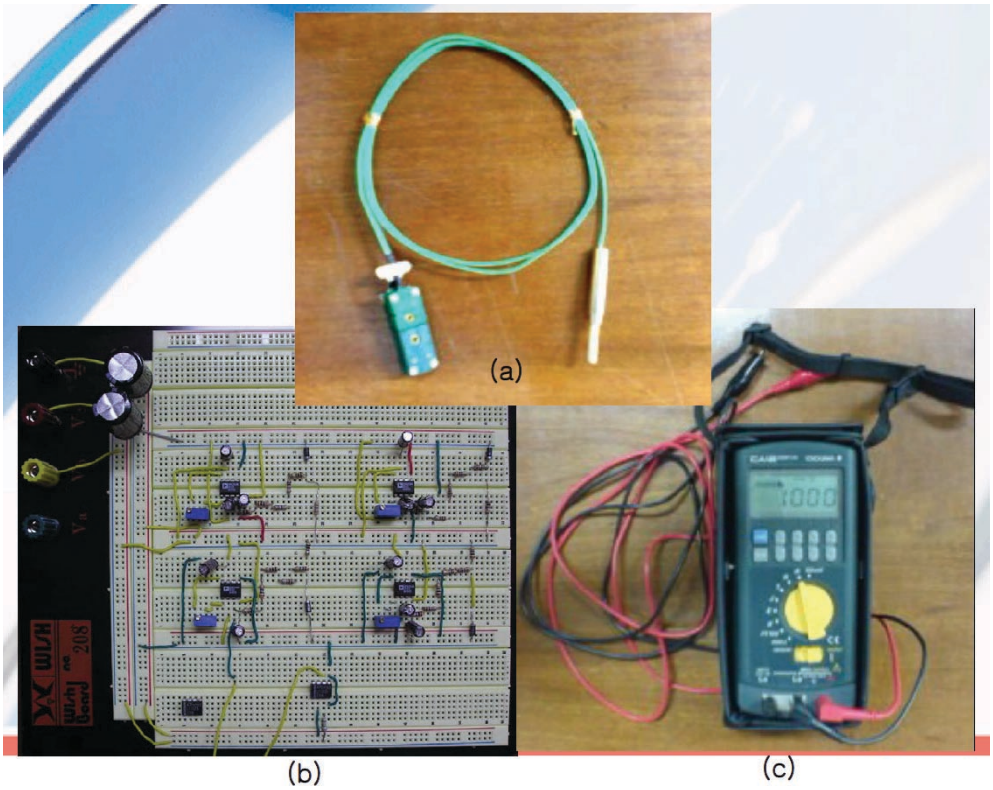


Figure 11. Experimental environment: (a) R-type TC, (b) test board, (c) TC simulator

In the cold compensation and disconnection detection circuit of various TCs in **Figure 6**, the measured values of the input voltage V_{IN} and output voltage V_{IOA} of the measurement amplifier for the steady-state and short-circuit detection circuits of various TCs (E, J, K, T, S-type) under the conditions in **Table 1** are shown in **Table 4**. In this table, the steady-state V_{IN} voltage is the electromotive force for each TC at room temperature (25°C) [9].

A summary of the characteristics of the conventional TC compensation and interface circuits [4,6-8] and the TC cold compensation and disconnection detection circuits proposed in this study is shown in **Table 5**. Compared with the conventional interfaces, the advantages of the interface limited in this study are 1) it is easy to implement TC cold compensation and disconnection detection for 8 channels, 2) the unit cost of interface construction is low, and 3) the cold point compensation is 97.8%.

Table 4. Input and output voltages of the instrumentation amplifier in case of disconnection and steady-state for various TCs

TC type	TC states			
	Normal		TC open	
	V_{IN} [mV]	V_{OIA} [V]	V_{IN} [mV]	V_{OIA} [V]
E	1.460	0.1899	-421.0	-14.215
J	1.278	0.1837	-420.5	-14.214
K	1.001	0.1825	-420.5	-14.215
T	0.993	0.4758	-421.0	-14.216
S	0.143	0.0768	-421.0	-14.215

Table 5. Summary of characteristics of conventional and proposed TC cold-junction compensation and disconnection detection circuits

Contents	Reference [4]	References [6–8]	This work
Supply voltage	-15 V – +15 V	2.7–5.5 V	-15 V – +15 V
Applicable TC type	R	K, J, T, N, S, E, B, R	K, J, T, S, E, R
Number of channels	1	1 (8 주 1)	8
Disconnection detection function	Yes	Yes (need to add circuits)	Yes
Unit price for o8-channel system configuration	LOW	High	LOW
Cold-junction compensation factor	97.8%	-	97.8%

8 주 1: 8 MCP9600 modules needed when configuring 8 channels with I^2C communication

Recently, due to the large capacity of memory, excellent linear characteristics of linear elements, and high speed of microprocessors, linearization of TCs is a trend to be processed by S/W. In this study, based on this, a program can be developed to store the electromotive force of TC according to temperature in memory, and then compare the analogue value measured from the TC interface circuit and the value stored in memory to determine the final linearized temperature value [13].

4. Conclusion

A circuit design with a thermocouple (TC) cold-junction compensation function and a disconnection detection function was developed, and a multi-channel interface circuit applicable to various TCs was implemented. When using an R-type TC, the error after cold-junction compensation of the TC was reduced by 97.8% from 0.14 mV to 3 μV. In addition, the output voltage of the measurement amplifier saturated from 88 mV to -14.2 V when the TC was disconnected from normal. In the designed multichannel interface circuit, only 4 channels were considered, and similar performance was obtained by applying and experimenting with R-type TCs. In addition, the cold-junction compensation and disconnection detection circuits and multichannel interface circuits of TCs limited in this study can be applied to E, J, K, T, and S-type TCs, and the conditions for use are presented. The results of this study will be very useful for the furnace for 300 mm wafers and the multichannel temperature control system using various TCs.

Disclosure statement

The author declares no conflict of interest.

References

[1] Carr JJ, 1993, Sensors and Circuits; Sensor Transducers, and Supporting Circuits for Electronic Instrumentation Measurement and Control, PTA Prentice Hall, USA.

[2] R-Type Thermocouple, n.d., http://www.sentech.co.kr

[3] Thermocouples, n.d., https://iseinc.com/Thermocouple.htm

[4] Cha HW, Kim YS, Park SH, et al., 2006, Design of R-Type Thermocouple Interface with Cold-Junction Compensator and its Broken Wire Detection, Proc. of IEEK Summer Conference, 847–848.

[5] LT1025 Datasheet, 2007, Linear Technology Co., https://www.datasheet.es/PDF/395136/LT1025-pdf.html

[6] MCP960X/L0X/RL0X Datasheet, Thermocouple EMF to Temperature Converter, ± 1.5°C Maximum Accuracy, 2015–2021, Microchip Technology Inc., https://ww1.microchip.com/downloads/en/DeviceDoc/MCP960X-L0X-RL0X-Data-Sheet-20005426F.pdf

[7] Adafruit MCP9600 I2C Thermocouple Amplifier, 2023, https://www.adafruit.com/product/4101

[8] Thermocouple Temperature Sensor, n.d., https://github.com/Seeed-Studio/Seeed_MCP9600

[9] Type R Thermocouple, n.d., https://www.thermocoupleinfo.com/type-r-thermocouple.htm

[10] AD620 Datasheet, 2007, Analog Device Co., https://www.analog.com/en/products/ad620.html

[11] Process Calibrator CA300 series, Source and Measure Simplicity, n.d., YOKOGAWA, https://cdn.tmi.yokogawa.com/BUCA300-EN.pdf

[12] User's Manual, CA310 Volt mA Calibrator, CA 320 TC Calibrator, CA330 RTD Calibrator Getting Started Guide, 2023, YOKOGAWA, https://web-material3.yokogawa.com/IMCA310-02EN.pdf

[13] Kim JL, 2004, Implementation of the Temperature Control System Using K-type Thermocouple. Journal of the Korea Society of Computer and Information, 9(3): 127–133.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

A Study on the Design Method of Hybrid MOSFET-CNTFET Based SRAM – A Secondary Publication

Geunho Cho*

Department of Electronic Engineering, Seokyeong University, Seoul, South Korea

***Corresponding author:** Geunho Cho, choroot@skuniv.ac.kr

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: More than 10,000 carbon nanotube field-effect transistors (CNTFETs) have been successfully integrated into one semiconductor chip using conventional semiconductor design procedures and manufacturing processes. These transistors offer advantages such as high carrier mobility, large saturation velocity, low intrinsic capacitance, flexibility, and transparency. The three-dimensional multilayer structure of the CNTFET semiconductor chip, along with ongoing research in CNTFET manufacturing processes, increases the potential for creating a hybrid MOSFET-CNTFET semiconductor chip. This chip combines conventional metal-oxide-semiconductor field-effect transistors (MOSFETs) and CNTFETs in one integrated system. This paper discusses a methodology to design 6T binary static random-access memory (SRAM) using a hybrid MOSFET-CNTFET. This paper introduces a method for designing a hybrid MOSFET-CNTFET SRAM by leveraging existing MOSFET SRAM or CNTFET SRAM design approaches. Additionally, this paper compares its performance with conventional MOSFET SRAM and CNTFET SRAM designs.

Keywords: MOSFET; CNTFET; SRAM; Hybrid; Carbon nanotube

Online publication: January 29, 2024

1. Introduction

For a long time, integrated circuits have increased the speed of circuits, lowered operating voltages, and improved integration by scaling down semiconductor device dimensions. However, as the channel length of metal-oxide-semiconductor field-effect transistors (MOSFETs) has decreased to the nanometer scale, additional improvements in MOSFET performance have become difficult due to short channel effects such as drain-induced barrier lowering (DIBL), gate-induced drain leakage (GIDL), hot carrier effects, punch through, and leakage current. While the introduction of fin-shaped field-effect transistors (FinFETs) has helped alleviate some of these issues, there is ongoing demand in the market for continuous performance enhancement of semiconductor devices, driving interest in new materials and structures to further improve the performance of existing semiconductor devices [1-6].

Connecting the source and drain of conventional semiconductor devices with highly conductive carbon nanotubes (CNTs) has enhanced the performance of these devices, overcoming many known issues

and limitations, and positioning them as promising next-generation semiconductor devices. CNTs used in carbon nanotube field-effect transistors (CNTFETs) are broadly classified into semiconducting CNTs and metallic CNTs, with the need for techniques to extract only semiconducting CNTs to control the current in CNTFETs. Currently, techniques for extracting only semiconducting CNTs have achieved purity levels of up to 99.99%. Furthermore, the distribution of CNT diameters, which determines the threshold voltage distribution of CNTFETs, has been steadily decreasing, contributing to the realization of stable CNTFET circuit implementations. In addition, physical and chemical process improvements of CNTFETs or CNTs are expected to play a crucial role in enhancing the performance of CNTFETs and expanding their range of applications. These improvements will enhance parameters such as electron mobility, on/off ratio, operating frequency, power consumption, performance differences between N-channel FETs (N-FETs) and P-channel FETs (P-FETs), and substrate flexibility, thus improving the performance of CNTFETs and expanding their utility [1-12].

CNTFETs with these advantages have raised expectations for the mass production of semiconductor chips utilizing CNTFETs, using existing industrial standard design procedures and process technologies to integrate more than 10,000 CNTFETs on a single chip. Moreover, the manufactured CNTFET chip, as shown in **Figure 1**, is structured in a three-dimensional multilayer design using existing semiconductor processes, demonstrating the possibility of implementing MOSFETs on one layer and CNTFETs on another. This suggests the potential for manufacturing hybrid MOSFET-CNTFET chips using existing semiconductor processes. Additionally, while most CNTFETs have been fabricated by placing CNTs on a silicon wafer (**Figure 2a**) and stacking drain or source contacts on top, recently developed inkjet printing technologies, as depicted in **Figure 2b**, allow for precise placement of CNTs on structures similar to conventional MOSFETs, opening up various possibilities for manufacturing hybrid MOSFET-CNTFET chips [5,6,13,14].

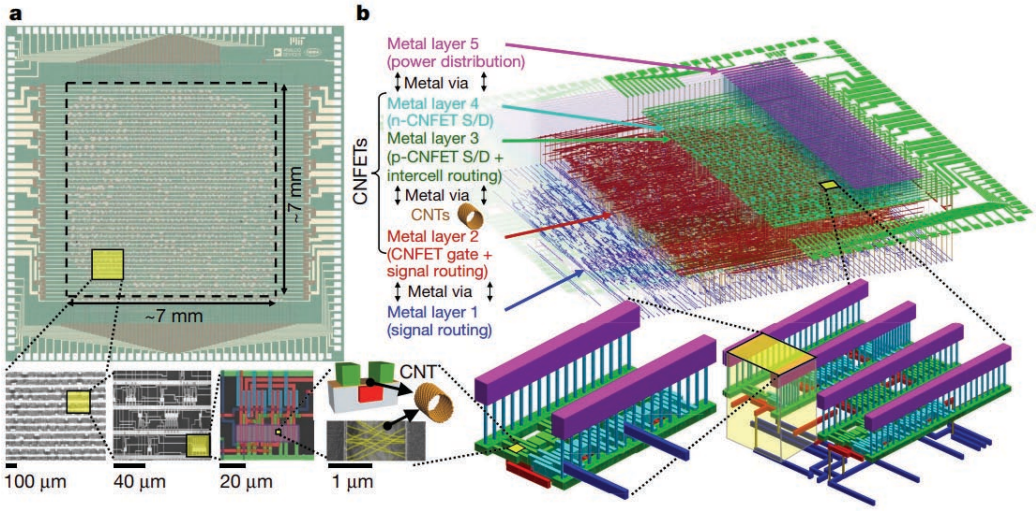


Figure 1. Microprocessor built from CNTFET [13]

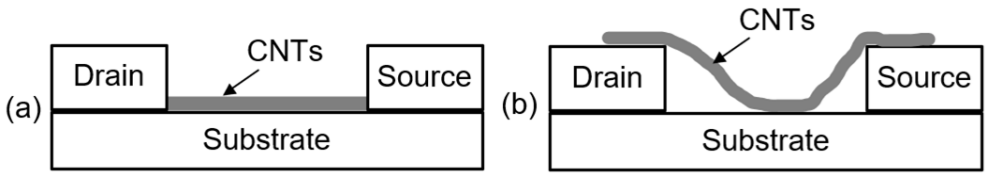


Figure 2. (a) Conventional deposition of CNTs [5]; **(b)** Deposition of CNTs by inkjet-printing [14]

In computer systems, static random-access memory (SRAM) is a crucial circuit block that needs to provide data for process operations at high speed. Typically, it occupies around 70% of the semiconductor chip area and is composed of SRAM cells that store minimal unit data such as 0 or 1, arranged in a repetitive pattern inside the chip. The performance of SRAM directly impacts the performance of microprocessors. Therefore, techniques to increase the speed of SRAM, reduce power consumption, and handle more data (ternary or quaternary) reliably have been continuously developed. Recently, various SRAM design studies utilizing CNTFETs have been underway [15-20].

Based on the current technological trends, the most realistic approach to implementing CNTFET chips at present involves designing circuit blocks as simple repetitive structures across the entire semiconductor chip. This reduces the impact of CNTFET process variations, allowing the implementation of circuit blocks with higher performance using CNTFET circuits compared to those implemented only with MOSFETs. In previous research [21], methods for designing SRAM with CNTFET were discussed, and their performance was compared with conventional MOSFET SRAM. This paper takes a step further by proposing a design method for hybrid SRAM, where complex data storage parts of SRAM cells are implemented with MOSFETs, and the data interconnection parts linking data storage and bit lines are composed of CNTFETs. The performance of the proposed hybrid MOSTFET-CNTFET SRAM will be compared with that of conventional MOSFET SRAM and CNTFET SRAM to assess the degree of performance improvement. For simulation verification, this paper utilizes the 32 nm PTM MOSFET library file and the Stanford 32 nm CNFET library file [22,23]. The aim is to discuss the design method for hybrid MOSFET-CNTFET SRAM using the existing design methods proposed in previous research for MOSFET SRAM or CNTFET SRAM and to compare their performance.

2. Hybrid MOSFET-CNTFET SRAM

This paper aims to discuss the design and performance enhancement of hybrid MOSFET-CNTFET SRAM based on the most fundamental SRAM cell known as the 6T SRAM cell. Referring to previous research results [21], it is referred to as hybrid SRAM for the sake of discussion efficiency. Hybrid SRAM, as shown in **Figure 3**, consists of a data storage section (MN3, MN4, MP5, and MP6) and a section (MN1 and MN2) connecting it to the bit lines (BL and BLB). This paper proposes a hybrid SRAM design that is less sensitive to process variations by constructing the relatively complex data storage section with MOSFETs, and the section connected to the bit lines with CNTFETs. In **Figure 3**, the transistors comprising the hybrid SRAM are represented using widely known conventional symbols for MOSFETs and include cylindrical symbols representing CNTs inside the CNTFETs to distinguish them from conventional MOSFET symbols. Furthermore, the channel of each transistor is depicted in grey for N-FETs and white for P-FETs.

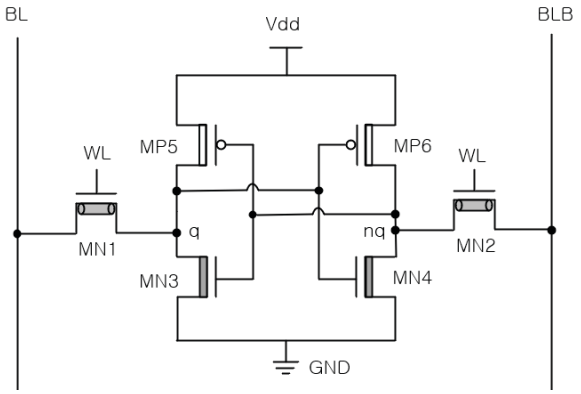


Figure 3. Hybrid MOSFET-CNTFET SRAM cell

To design the hybrid SRAM, it is necessary to determine the values of MN4/MN2 and MP5/MN1, similar to the conventional 6T SRAM design method. The “nq” value should be designed to be smaller than the threshold voltage of semiconductor devices to ensure that the stored data in the SRAM cell remains unchanged during reading. Since the data storage section in this paper is composed of MOSFETs, a threshold voltage of 0.18 V for MOSFETs was considered when considering the threshold voltage. The voltage value of “nq” according to MN4/MN2 is depicted in **Figure 4**, taking into account the CNT density of the CNTFETs when discussing each MN4/MN2. Specifically, CNT1 indicates that one CNT is placed within a 32 nm gate width, and CNT3 indicates that three CNTs are evenly spaced within the gate width. Similarly, the voltage value of “q” according to the ratio of MP5/MN1 and the CNT density is shown in **Figure 5**, and the “q” voltage value should be selected lower than the threshold voltage of MOSFETs to enable writing new data to the SRAM cell.

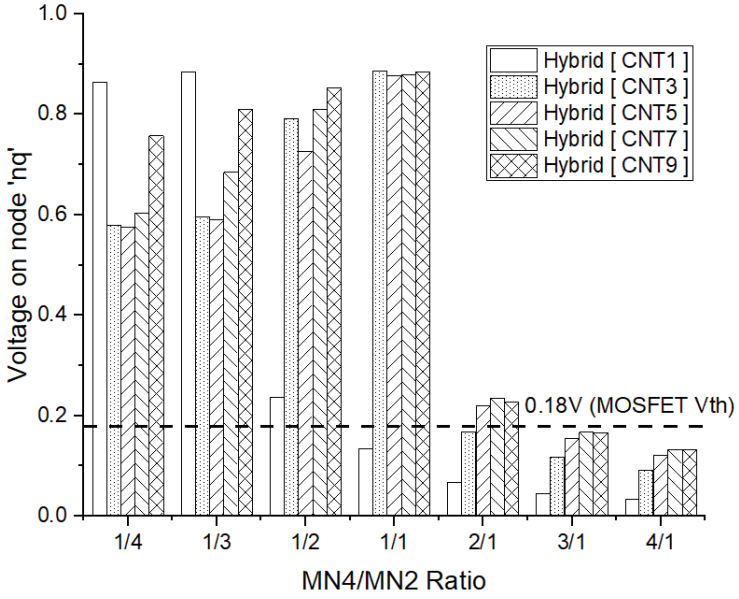


Figure 4. Voltage on node “nq” depending on MN4/MN2 ratio

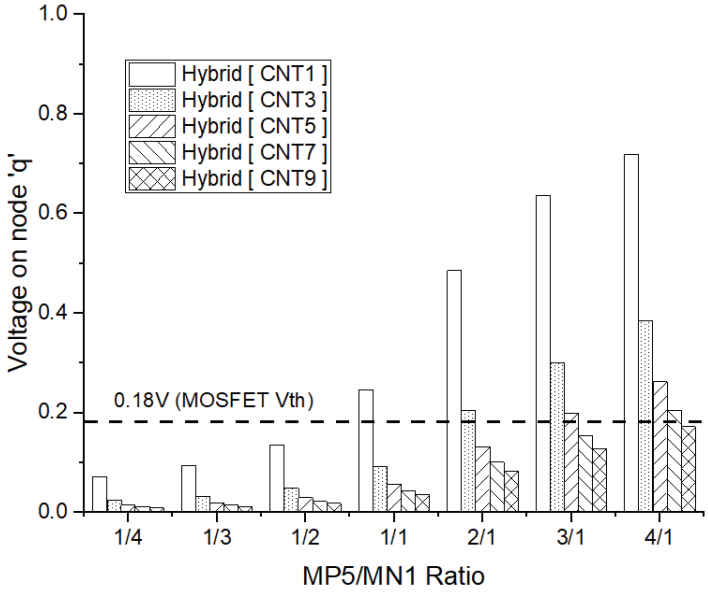


Figure 5. Voltage on node “q” depending on MP5/MN1 ratio

Through **Figures 4** and **5**, it was possible to use CNT7 and CNT9 when the ratios of MN4/MN2 and MP5/MN1 were 3/1, and only CNT9 could be used when the ratios were 4/1. In this paper, to minimize the area of the SRAM cell and reduce the possibility of process defects due to CNT arrangement, the ratios of MN4/MN2 and MP5/MN were selected as 3/1, and the CNT density was chosen as CNT7. Subsequently, the performance was compared with conventional SRAMs, i.e. all MOSFET SRAMs and all CNTFET SRAMs.

Table 1 shows the performance differences between all MOSFET SRAM, all CNTFET SRAM, and hybrid SRAM. In terms of read delay, hybrid SRAMs show performance comparable to all CNTFET SRAM and higher than all MOSFET SRAM. However, compared to all CNTFET SRAM, there is not much performance improvement, and instead, it consumes about 20 times more power. This indicates that CNTFETs can pass more current per unit area than MOSFETs, resulting in unnecessarily high current flow during read operations. This phenomenon can also be observed in the voltage characteristics of the "nq" node in **Figure 4**. In **Figure 4**, when MN4/MN2 is 1/4, 1/3, 1/2, and 1/1, it can be seen that if MN4 cannot pass enough current due to the significant current flow of MN2, the "nq" node does not decrease sufficiently in size to perform read operations easily. However, in write operations, such hybrid SRAM contributes to the performance enhancement of conventional SRAM. In write operations, all CNTFET SRAM does not show significantly better performance than all MOSFET SRAM and consumes more than twice the power. However, hybrid SRAM shows a delay reduction of more than three times compared to all MOSFET SRAM at similar power consumption. In other words, in terms of power delay product (PDP), all CNTFET SRAM improves read performance compared to all MOSFET SRAM but worsens write performance. However, hybrid SRAM overall enhances both read and write performance compared to all MOSFET SRAM. Therefore, unless high performance in read operations is crucial, it is unnecessary to manufacture SRAM with complex data storage sections using CNTFETs, even at the expense of higher process variation risks.

Table 1. Delay, power, and power delay product (PDP) of all MOSFET SRAM, all CNTFET SRAM, and hybrid SRAM

	Operation	CNT density	Delay (ps)	Power (μW)	PDP (aJ)
All MOSFET SRAM [23]	Read	N/A	234.90	0.041	9.54
	Write		35.58	0.604	21.48
All CNTFET SRAM [23]	Read	7	116.00	0.004	1.00
	Write		33.82	1.407	47.58
Hybrid SRAM	Read	7	104.00	0.080	8.31
	Write		11.70	0.752	8.80

3. Conclusion

In the era of the Internet of Things (IoT), the expectations for CNTFETs are steadily increasing, as they offer higher performance than conventional semiconductor devices, enabling not only data processing but also various applications in areas such as general-purpose sensors and wearable healthcare devices. Although mass production technologies for CNTFET-based semiconductor chips have not yet matured enough for commercialization, the utilization of CNTFETs even some circuit blocks of overall semiconductor chips could significantly enhance the feasibility of implementing CNTFET semiconductor chips. Therefore, this paper proposed a method for designing a hybrid SRAM composed of MOSFETs and CNTFETs, demonstrating that it achieves over three times faster write operations and twice the power improvement compared to SRAMs

composed solely of CNTFETs. This suggests a promising avenue for further exploration and advancement in the integration of CNTFETs into semiconductor chip designs.

Acknowledgment

This work was supported by Seokyeong University in 2022. The EDA tool was supported by the IC Design Education Center (IDEC), Korea.

Disclosure statement

The author declares no conflict of interest.

References

[1] Franklin AD, Hersam MC, Wong H-SP, 2022, Carbon Nanotube Transistors: Making Electronics from Molecules. Science, 378(6621): 726–732. https://doi.org/10.1126/science.abp8278

[2] Zhang S, Pang J, Li Y, et al., 2022, Emerging Internet of Things Driven Carbon Nanotubes-Based Devices. Nano Research, 15: 4613–4637. https://doi.org/10.1007/s12274-021-3986-7

[3] Chen R, Chen L, Liang J, et al., 2022, Carbon Nanotube SRAM in 5-nm Technology Node Design, Optimization, and Performance Evaluation – Part I: CNFET Transistor Optimization. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 30(4): 432–439. https://doi.org/10.1109/TVLSI.2022.3146125

[4] Chen R, Chen L, Liang J, et al., 2022, Carbon Nanotube SRAM in 5-nm Technology Node Design, Optimization, and Performance Evaluation – Part II: CNT Interconnect Optimization. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 30(4): 440–448. https://doi.org/10.1109/TVLSI.2022.3146064

[5] Deng J, Wong H-SP, 2007, A Compact SPICE Model for Carbon-Nanotube Field-Effect Transistors Including Nonidealities and Its Application – Part I: Model of the Intrinsic Channel Region. IEEE Transactions on Electron Devices, 54(12): 3186–3194. https://doi.org/10.1109/TED.2007.909030

[6] Deng J, Wong H-SP, 2007, A Compact SPICE Model for Carbon-Nanotube Field-Effect Transistors Including Nonidealities and Its Application – Part II: Full Device Model and Circuit Performance Benchmarking. IEEE Transactions on Electron Devices, 54(12): 3195–3205. https://doi.org/10.1109/TED.2007.909043

[7] Yang D, Hwang H, Kim Y-J, et al., 2023, High-Performance Carbon Nanotube Field-Effect Transistors with Electron Mobility of 39.4 $cm^2V^{-1}s^{-1}$ Using Anion–π Interaction Doping. Carbon, 203: 761–769. https://doi.org/10.1016/j.carbon.2022.12.025

[8] Zhou J, Ren L, Li H, et al., 2023, Carbon Nanotube Radiofrequency Transistors With f_T/f_{MAX} of 376/318 GHz. IEEE Electron Device Letters, 44(2): 329–332. https://doi.org/10.1109/LED.2022.3227133

[9] Liu C, Cao Y, Wang B, et al., 2022, Complementary Transistors Based on Aligned Semiconducting Carbon Nanotube Arrays. ACS Nano, 16(12): 21482–21490. https://doi.org/10.1021/acsnano.2c10007

[10] Long G, Jin W, Xia F, et al., 2022, Carbon Nanotube-Based Flexible High-Speed Circuits with Sub-Nanosecond Stage Delays. Nature Communications, 13(1): 6734. https://doi.org/10.1038/s41467-022-34621-x

[11] Shin D-H, You YG, Jo SI, et al., 2022, Low-Power Complementary Inverter Based on Graphene/Carbon-Nanotube and Graphene/MoS_2 Barristors. Nanomaterials, 12(21): 3820. https://doi.org/10.3390/nano12213820

[12] Huang Q, Wang J, Li C, et al., 2022, Intrinsically Flexible All Carbon-Nanotube Electronics Enabled By a Hybrid Organic-Inorganic Gate Dielectric. npj Flexible Electron, 6: 61. https://doi.org/10.1038/s41528-022-00190-8

[13] Hills G, Lau C, Wright A, et al., 2019, Modern Microprocessor Built From Complementary Carbon Nanotube

Transistors. Nature, 572: 595–602. https://doi.org/10.1038/s41586-019-1493-8

[14] Singh RS, Takagi K, Aoki T, et al., 2022, Precise Deposition of Carbon Nanotube Bundles by Inkjet-Printing on a CMOS-Compatible Platform. Materials, 15(14): 4935. https://doi.org/10.3390/ma15144935

[15] Valluri A, Musala S, 2024, Low Leakage, Differential Read Scheme CNTFET Based 9T SRAM Cells for Low Power Applications. International Journal of Electronics, 111(1): 127–148. https://doi.org/10.1080/00207217.2022.2148289

[16] Bakhtiary V, Amirany A, Moaiyeri MH, et al., 2023, An SEU-Hardened Ternary SRAM Design Based on Efficient Ternary C-Elements Using CNTFET Technology. Microelectronics Reliability, 140: 114881. https://doi.org/10.1016/j.microrel.2022.114881

[17] Sachdeva A, Kumar D, Abbasian E, 2023, A Carbon Nano-Tube Field Effect Transistor Based Stable, Low-Power 8T Static Random Access Memory Cell with Improved Write Access Time. AEU – International Journal of Electronics and Communications, 162: 154565. https://doi.org/10.1016/j.aeue.2023.154565

[18] Darabi A, Salehi MR, Abiri E, 2023, One-Sided 10T Static-Random Access Memory Cell for Energy-Efficient and Noise-Immune Internet of Things Applications. International Journal of Circuit and Theory Applications, 51(1): 379–397. https://doi.org/10.1002/cta.3408

[19] Elangovan M, Muthukrishnan M, 2022, Design of High Stability and Low Power 7T SRAM Cell in 32-NM CNTFET Technology. Journal of Circuits, Systems and Computers, 31(13): 2250233. https://doi.org/10.1142/S0218126622502334

[20] Bastani NH, Navi K, 2022, A Low-Power and Robust Quaternary SRAM Cell for Nanoelectronics. Analog Integrated Circuits and Signal Processing, 111: 483–493. https://doi.org/10.1007/s10470-022-02031-0

[21] Cho G, 2021, A Study on the Circuit Design Method of CNTFET SRAM Considering Carbon Nanotube Density. Journal of IKEEE, 25(3): 473–478. https://doi.org/10.7471/ikeee.2021.25.3.473

[22] PTM Models, viewed February 1, 2023. http://ptm.asu.edu/

[23] CNFET Models, viewed February 1, 2023. https://nano.stanford.edu/downloads/stanford-cnfet-model/stanford-cnfetmodel-hspice

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Integrated Services Platform of International Scientific Cooperation

Innoscience Research (Malaysia), which is global market oriented, was founded in 2016.Innoscience Research focuses on services based on scientific research. By cooperating with universities and scientific institutes all over the world, it performs medical researches to benefit human beings and promotes the interdisciplinary and international exchanges among researchers.

Innoscience Research covers biology, chemistry, physics and many other disciplines. It mainly focuses on the improvement of human health. It aims to promote the cooperation, exploration and exchange among researchers from different countries. By establishing platforms, Innoscience integrates the demands from different fields to realize the combination of clinical research and basic research and to accelerate and deepen the international scientific cooperation.

Cooperation Mode



Clinical Workers



In-service Doctors





Foreign Researchers





Hospital



University



Scientific institutions

BIO-BYWORD
SCIENTIFIC PUBLISHING PTY LTD

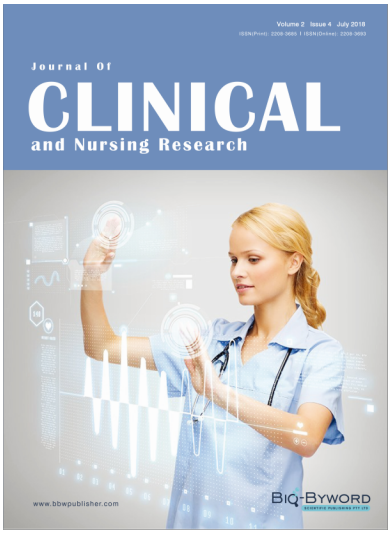
OUR JOURNALS



The *Journal of Architectural Research and Development* is an international peer-reviewed and open access journal which is devoted to establish a bridge between theory and practice in the fields of architectural and design research, urban planning and built environment research.

Topics covered but not limited to:

- Architectural design
- Architectural technology, including new technologies and energy saving technologies
- Architectural practice
- Urban planning
- Impacts of architecture on environment



Journal of Clinical and Nursing Research (JCNR) is an international, peer reviewed and open access journal that seeks to promote the development and exchange of knowledge which is directly relevant to all clinical and nursing research and practice. Articles which explore the meaning, prevention, treatment, outcome and impact of a high standard clinical and nursing practice and discipline are encouraged to be submitted as original article, review, case report, short communication and letters.

Topics covered by not limited to:

- Development of clinical and nursing research, evaluation, evidence-based practice and scientific enquiry
- Patients and family experiences of health care
- Clinical and nursing research to enhance patient safety and reduce harm to patients
- Ethics
- Clinical and Nursing history
- Medicine



Journal of Electronic Research and Application is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application.

Topics covered but not limited to:

- Automation
- Circuit Analysis and Application
- Electric and Electronic Measurement Systems
- Electrical Engineering
- Electronic Materials
- Electronics and Communications Engineering
- Power Systems and Power Electronics
- Signal Processing
- Telecommunications Engineering
- Wireless and Mobile Communication