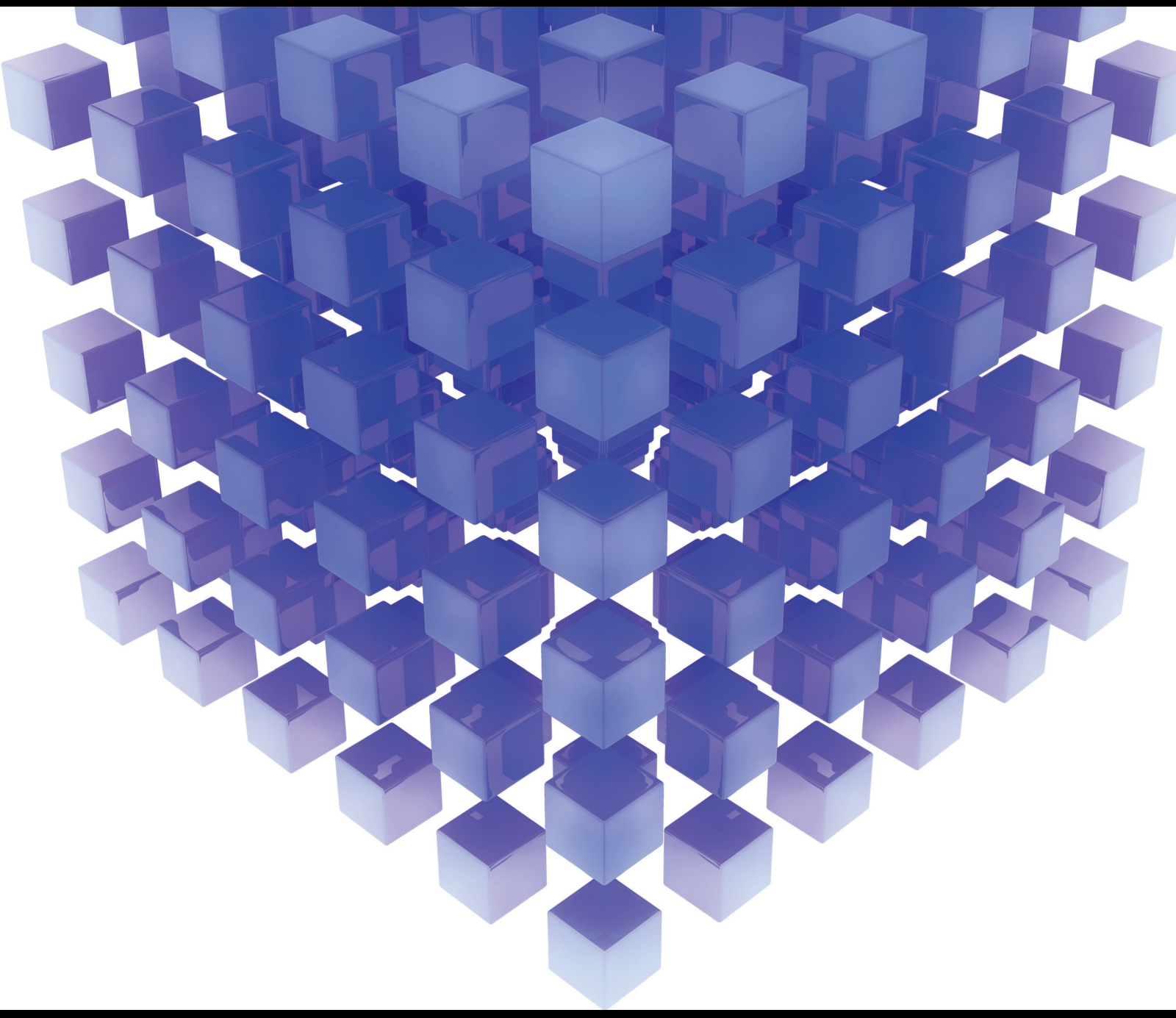
Mathematical Problems in Engineering

Imperfect Real Data Handling in an Internet of Things Environment

Lead Guest Editor: Xianyong Li
Guest Editors: Mu Zhou, Nan Jiang, and Xiaobin Zhu



Imperfect Real Data Handling in an Internet of Things Environment

Mathematical Problems in Engineering

Imperfect Real Data Handling in an Internet of Things Environment

Lead Guest Editor: Xianyong Li
Guest Editors: Mu Zhou, Nan Jiang, and Xiaobin Zhu

Chief Editor

Academic Editors

Contents

Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 9835797, 1 page
https://doi.org/10.1155/2023/9835797

Retraction

Retracted: University Piano Education Visualization System under the Background of Distance Education Based on 5G Network

Mathematical Problems in Engineering

Received 17 October 2023; Accepted 17 October 2023; Published 18 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Xie, "University Piano Education Visualization System under the Background of Distance Education Based on 5G Network," *Mathematical Problems in Engineering*, vol. 2022, Article ID 6825591, 7 pages, 2022.

Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 9835797, 1 page
https://doi.org/10.1155/2023/9835797

Retraction

Retracted: University Piano Education Visualization System under the Background of Distance Education Based on 5G Network

Mathematical Problems in Engineering

Received 17 October 2023; Accepted 17 October 2023; Published 18 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Xie, "University Piano Education Visualization System under the Background of Distance Education Based on 5G Network," *Mathematical Problems in Engineering*, vol. 2022, Article ID 6825591, 7 pages, 2022.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 6825591, 7 pages
https://doi.org/10.1155/2022/6825591

Research Article

University Piano Education Visualization System under the Background of Distance Education Based on 5G Network

Shali Xie

School of Architecture and Art, Central South University, Changsha, Hunan, China

Correspondence should be addressed to Shali Xie; 2111602008@e.gzhu.edu.cn

Received 30 August 2021; Revised 7 October 2021; Accepted 15 October 2021; Published 9 March 2022

Academic Editor: Xianyong Li

University piano education online learning courses supplement learning and have quickly advanced distance education and the essence of improving Internet innovation. Experts of piano education visualization system and distance piano education visualization system are a world specification, which offers various university and online open-source courses, for those who seek an online expert for piano education events. The individual online piano learning course is another choice for learning the education visualization system via progressive Internet correspondence. Past techniques depend on the profound learning piano education system courses of learning and information mining distance education. Existing strategies present specialized difficulties, which include the difficulty of the piano learning via an online network. As such, this study aims to establish a university piano education visualization system framework via a distance education background utilizing the fifth-generation (5G) network; the distance education learning for piano education visualization system course using the 5G network is proposed. The remote sensor shows an assortment of methods for grownups to partake in the online learning preparation. The idea of online communication within a network of instructors and understudies is extraordinary. When considering training-based or execution-based courses, there is a likelihood that the learning system may exhibit an inferior quality of sound, and the sound of the piano notes may lag with time over the network connection. Online learning provides adaptability and internationalization and has the capacity to interface with countless individuals and bring them together in the online learning environment. In this situation, the straightforward transportation and installment of remote sensors, the user of the distance educator 5G network is advised to view it from a commonsense perspective. Furthermore, there is a need to advance online piano education training in instructive exercises.

1. Introduction

Piano education visualization system of learning may very well be one, which includes the structure and related information. The fundamental learning piano education visualization system portrays a combination of innovation for online correspondence that has been conducted in learning history. Structure, exercises, assessment, and other major useful aspects act as an illustration of the current practice. Educational program configuration specialists, who utilized instruction using the word processor in the learning environment, presented a specialized review during the piano courses' advancement. This distance education learning features provided additional improvement in new learning style. To decide the online courses' idea, we should change the significant capacity of instruction practice and the online learning educator's online learning educator, learning the straight-line distance education system's training. The current arrangement is that the piano education visualization system practice in the piano educational climate is previously a model of distance piano education visualization system. These models are wide-spread and have greatly influenced the students through online training, which is a requirement for recorded proof, advancement content, and distance instructing. The model also contains a conveyance strategy that the understudy has been endorsed for the undergrad and graduate online courses and Internet learning. By and large, these models have been utilized conversely with "on the web" and "distance education." Increasingly more the utilization of distance learning "on the web" has become an assistance understudy to online learning course. A professional presents the endorsement's

learning business or arranged to improve their learning training aptitudes in the course. The program, which has a cutoff time for finishing a predetermined course, will have the option to download talks and materials as needed.

The most recent working framework associated with the fast Internet and can download uncommon programming. The machine learning algorithm will get familiar with the learning courses, for example, aptitudes and learning treatment of learning. The online learning course covers numerous preparing learning territories, which include examining the learning hypotheses, directing, and piano learning. In certain focuses, the understudy says the courses should be taken in a degree program. The piano learning course generally takes an entire semester, and the 5G network can make beneficial contributions to degree courses. For instance, a learning program for the training project of the current learning instruction of understudies and their folks, piano learning educators, directors, and partners.

Figure 1 shows the assessment process in government-funded schools. Understudies who would not usually opt to attend world-class courses are provided with the opportunity to assess numerous parts of the arrangement. Understudies joining the program will gain the substantial advantages of social and scholarly projects, such as learning training.

Undergraduates appear to have already partaken in this program. For instance, when learning the piano education visualization system, partners started to estimate a learning program. When the performers and understudies of learning training discover what different artists are doing themselves, they become familiar with the information independently of others. The attitudes of many youthful understudies could be changed by challenging this learning program.

When preparing the courses of an online piano education, the difficult work of preparing the understudies is the final execution of the year. Students can familiarize themselves with their aptitudes inside one month. The activities assigned to understudies should not overwhelm students, but provide them with gradually increasing levels of skill sets.

The remainder of this article is arranged as follows. Section 2 reviews the literature on piano education visualization systems. Section 3 constructs the proposed architecture-based distance education program based on the 5G network. Section 4 performs the validation and verification analyses of the piano education visualization system. Section 5 concludes the paper.

2. Literature Survey

Distance training and distance instruction education have greatly advanced in recent years. In remote piano-education visualization systems, planning and improvement are essential for assuring adequate instruction. The current educational plan of piano-education visualization systems has several downsides. After examining the program, a portion of the proposals are selected for enhancement and developmental improvements [1]. Piano education provides online trials to understudies. Staff at a remote Spanish university

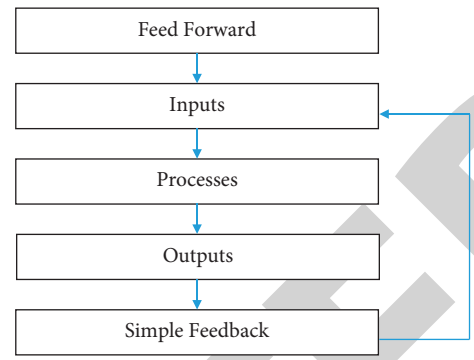


Figure 1: Block diagram of the wireless sensor.

have already begun the planning and execution of distant laboratories. This article focuses on the planning of a piano education visualization system and the advancement of electronic equipment (such as user and control devices) at a remote research center [2]. New data and correspondence advances have ushered in a new age of distance instruction. Development of present innovations will provide cutting-edge conveyance devices for transmitting distant information. Lord Faisal University has begun a distance learning program using an adaptable model for educating and learning. Simultaneous and offbeat communications have created a favorable environment for instruction [3]. Understudies in remote laboratories cannot access conventional research centers, but can complete their online tests and assignments at any time and place using a virtual homeroom instructor. The primary downside of such a research facility is its energy cost. Toward a sustainable, environmentally friendly remote-learning system, this work proposes the harnessing of energy from the sun, wind, and other renewables [4]. During continuous correspondence with educators, understudies can perform online distance-education visualization exercises with guaranteed input from instructors. Meeting the continuous necessities of the electronic whiteboard has bottlenecked the further improvement of online distance instruction. A strategy dependent on multimedia streaming innovation is required [5].

Versatile learning has become a new learning model [6, 7]. Finding additional opportunities for instruction is one goal of versatile innovation research. Dissecting the exploration circumstance and advancement pattern of versatile innovation and presenting the ramifications of portable learning, the authors of [8] proposed a hypothetical centrality of current portable distance instruction [8]. The outcomes show that the appropriateness and vital arrangements exercises of the quality prerequisites of the instruction idea of actual advanced water of the window ornament, the consciousness of the issues identified with the updates and direction abilities are absent. These outcomes are also used to pursue poll information apparatuses for investigating the significance of computerized screen-quality standard model outcomes and the significance of an advanced education setup [9]. Based on the above discussion, a community-oriented distance learning model of piano education would enable synergistic learning, point-by-point

conversations, sharing of references from piano-education training and investigation, acknowledgment of the application program, and the utilization of multimedia applications [10]. To examine the requirement for and achievability of building a unique remote piano-education visualization system framework that depends on distance instruction and preparation, we must compare the structure plans of distance learning frameworks [11]. With fast advances in science and innovation, distance training can assimilate an assortment of information and provide a graduation of stages to students. A portion of these challenges will test the mechanical plan of distance instruction. To improve the productivity and adequacy of instruction, training strategies for distant learning are required [12]. After conducting a survey, the authors of [9] proposed an information assortment device for a computerized picture-quality reference model that enhances the significance of advanced education. To objectively assess a data framework, the adequacy of present-day distance instruction is assessed in meetings and surveys [13]. New Internet tools such as the 5G network can be incorporated into sociologies. Students report that dynamic computerized pictures and information provide optimal restorative training [14, 15]. The upcoming 5G remote frameworks will require shaping by Internet-of-Things technologies. Web-of-Things innovations have begun changing the scenes of different ventures. Instruction strategies such as primary concerns, advancement of learning, and advice will be corresponded to students of the piano education visualization system [14, 16–18].

Although modern innovations promise to meet the necessities of 5G-based remote piano education, the network execution is fundamentally limited by the import co-channel impedance. The obstruction of executives is especially significant [19, 20]. A savvy piano education visualization system framework mirrors the arrangement of a smart city, utilizing the Internet of Things, the Internet, and current Internet-based intelligent application frameworks. To apply virtualization and clever innovation techniques, we require improved Internet services and recognizable proof through Radio Frequency Identification campus card technology [21].

3. Materials and Methods

Quicker is recognized as a 5G framework with adaptable remote usability. The 5G virtual network was the original cloud-based planning innovation. A specific type of 5G arrangement provides remote access to a central piano-education system. Contingent on the network design, the 4G or 5G network enables wireless or fixed connectivity. To provide 5G assistance in a wider setting, danger appraisals beyond monetary and administrative dangers might be required. Likewise, administrators can install powerful observing and control instruments that permit a previous admonition and react rapidly and successfully.

As shown in Figure 2, the piano-education stage of the proposed education scheme will identify the adaptable target level and set up the piano-education learning plan. The piano education and an objective degree of learning in the operational stage depend on the piano education's sending and tasks. The strategy for planning the learning identified with these of the common are substantial. Connection over the 5G network assumes a degree of trust between the entertainer and the identifier.

3.1. Distance Education Based on the 5G Network. Distance education training is typically offered as one course [22]. Online piano classes can be taken by students worldwide in their own spaces. Wireless sensors are principally utilized in the board programming of an online course. The fundamental motivation of an educational visualization system is to provide an advantageous learning system using the latest technologies.

Figure 3 is a block diagram of the remote learning system, showing the interactions between users, the domain, and the machine learning algorithm. Correspondence is a significant component of a circulated learning system. As learning difficulties and progresses are corresponded through dynamically distributed machines, they must be concentrated into a []. Instructors and learners alike can express their desires and exert impacts on the piano education visualization system, thus relaxing the constraints of traditional online learning services. In this way, partaken in the educator, the sincere belief of own, in two phases, including considering the distance piano education visualization system's job through experience and practice.

3.2. Data Extraction Using a Machine Learning Algorithm. The following outlines the steps of essential information-checking by artificial intelligence in the proposed system. Note that data on a two-dimensional worksheet or information-based table are hazy.

Step 1:

Dataset name: "Data Network."

Size of data set = 8.47 GB (9,096,733,606 bytes), 71,091,606 lines.

Step 2:

Number of cases in the relation (row) = 6640 URLs.

Number of features in the table = 3.

Step 3:

Attribute Description:

Id: definition number

Words: string

Step 4:

Hackers:{yes, no}

Step 5:

Upload dataset:

Upload the big data by pressing Preprocess and then on Open file. "Big Data-web hackers. Of ").

Step 6:

Decision tree procedure to "Big Data-web hackers.arff "

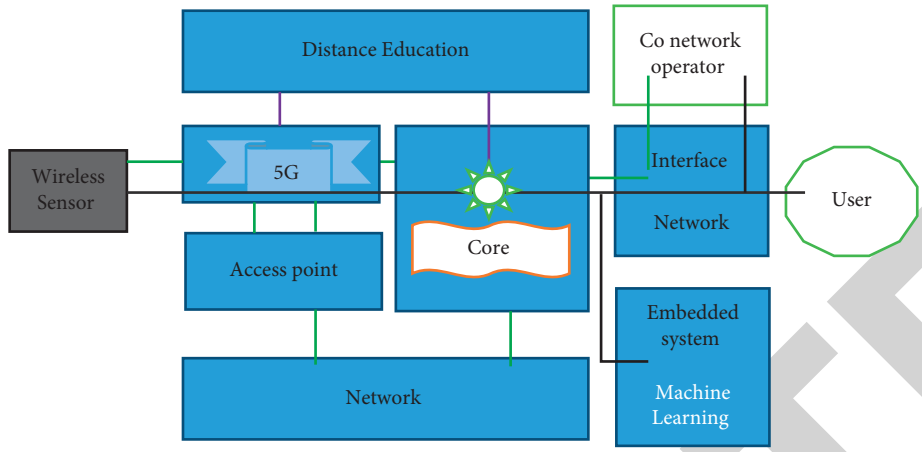


Figure: 2: Architecture of the proposed distance education system based on the 5G network.

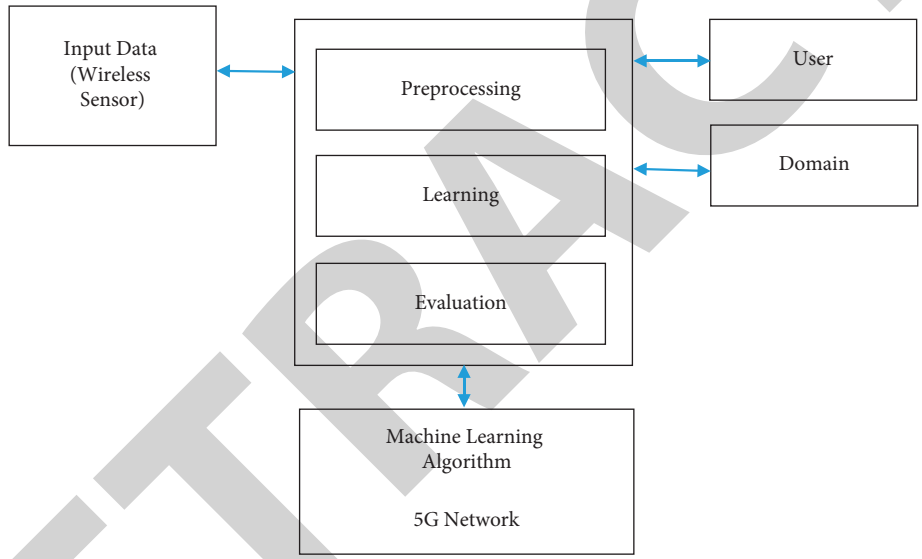


Figure 3: Block diagram of distance learning education course using a wireless sensor.

The learned data are immediately sent to the "delivered" procedures.

Step 7:

The above process returns the total number of operational data. In this case, the preparation adjusted 98% of the data (29,253 cases). These results cannot guarantee that the planned data will be enhanced by the results of the test set. More precisely, the result cannot infer the probability assessment of (0.1), as the root mean square (0.3) does not provide the average absolute difference equivalent. The clarification miss-step is not 0 or 1, implying that the model requires further refinement.

3.3. Wireless Sensor. Instructions delivered through remote sensors are (by definition) communicated among isolated understudies and instructors. Instructing diverse students from diverse places is much more challenging than instructing students in a standard study hall. To satisfy the needs and objectives of global remote learning, exhaustive remote sensors may be required. Understudies have expressed a need to communicate their feelings through the sensor network.

The wireless sensor network (Figure 4) allows students to express their views and feelings after completing their exercises. Likewise, they can gain input from the instructor via the remote sensor.

4. Results and Discussion

In execution-based online courses such as music, skill should not be assessed purely on the basis of understanding the material. To produce the correct sounds, music students should be instructed to communicate the development of their physical skill. This section discusses the benefits of the machine learning-based online piano-education learning plan.

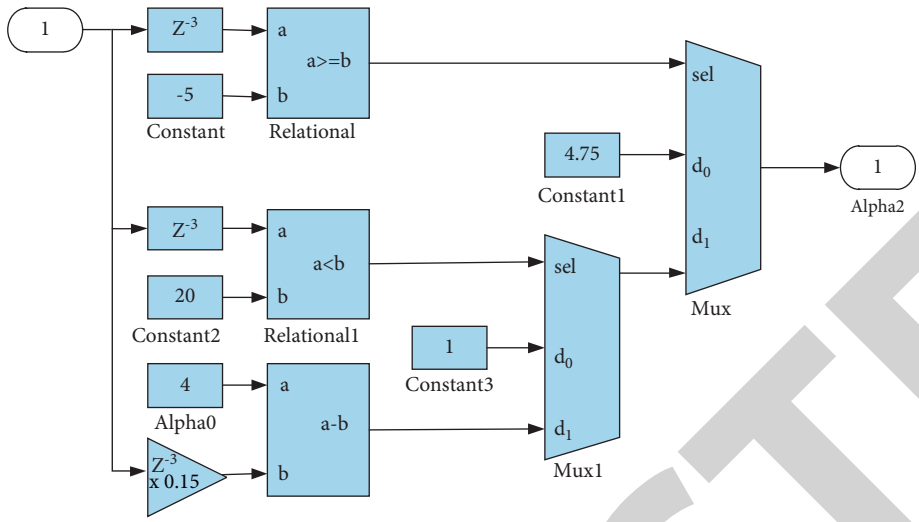


Figure 4: Circuit diagram of the wireless sensor network.

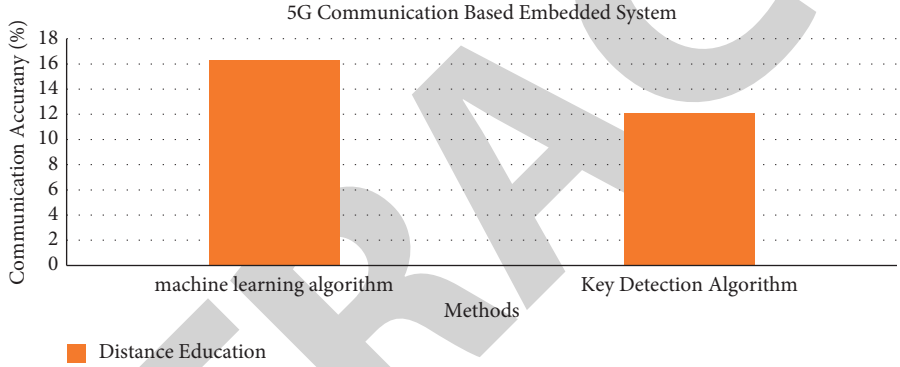


Figure 5: Comparison analysis of machine learning algorithm.

Table 1: 5G Network speeds of embedded [] in different periods.

No. of years	3G with embedded (Mbps)	4G with embedded (Mbps)	5G with embedded (Mbps)
2000–2004	200.1	256.1	350.1
2005–2010	275.1	290.1	360.2
2011–2015	279.1	295.1	400.1
2016–2020	283.1	292.1	450.1

Educators and instructors of web-based learning depict the focal objectives (educating and learning) from different points of view. Another advantage of the installed framework is the low power cost of the host framework.

The economy of a learning model largely depends on the capacities of the execution and answers provided. Considerable resources are invested in progressive tests for determining the precision of a model.

Figure 5 compares the communication accuracies of at least two segments of visual correspondence, which may include shapes, structures, images, and addition types, in the machine learning algorithm and the key detection algorithm. The strength of the 5G network was assessed by the Hesston model based on the trait choice-value. The Hesston model coefficients of market information needs were estimated by the least-squares strategy with nonlinear boundaries.

Table 1 shows the improvements in network speed over the years. Quantitative facilitating strategies, utilization of essential clearing plan setups, and improving the boundary alignment calculation will refine the model in future.

Improved hardware resources can also improve a substandard model. Connections between resources are freely available and resources compatible with the existing resources and current costs are continuously sought. The vanilla choice is expected to have been utilized as a count or a straightforward boundary alternative as an example.

Table 2 shows the outcomes of increasing the 5G size. Looking up completely and incredibly diminishing their expense is appropriated, i.e., more noteworthy than 5G executed it is consistently more prominent than the total enormous overhead time decreased 30%. In this situation, the information boundary is the realized product cost and

Table 2: 5G speedups in different asset configurations.

No. of underlying	Threads/5G (Mbps)	Cores/5G (Mbps)	Speedup/5G with embedded (Mbps)
4	4	36	350.1
8	3	26	202.2
16	2	20	129.1
32	2	12	123.1

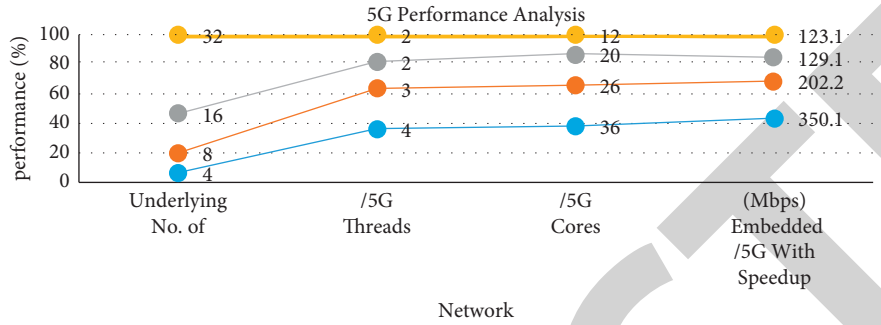


Figure 6: 5G performance analysis ratios.

the yield is obtained from various sources, which contrasts with the Hesston model.

Figure 6 shows the increases in 5G execution speed over the years.

5. Conclusion

Under appropriate guidance, online piano education can be completely taught online to a wide range of students. A remote music education system greatly contrasts with an on-campus music education, in which individuals meet on a daily basis. The 5G network provides a means of educating individuals who cannot meet the assigned timetable or are busy with other commitments. The focal connection of web-based learning and adaptability provides advance opportunities for professional improvement. Piano-sensing wireless sensors enable students to improve their skills or learn new ones, thereby expanding their business openings or simply providing them with individual satisfaction. Flexibility is a tremendous advantage of online learning, and the quality of the instructors is more important than online training per se when selecting a distance-based education system. Without diminishing the information and aptitude gains, online learning can be both economical and energy-saving.

In future studies, we will investigate the robustness of the university-level piano education visualization system based on the 5G network and its commercial value in practical applications. As the 6G network is the future development trend, we will also consider the possibility of upgrading the piano education visualization system to the 6G network.

Data Availability

Data are contained within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Hunan Province Philosophy and Social Science Foundation Project (batch: 16YBQ071).

References

[1] Y. Yang and D. Tian, "The research of multimedia curriculum design and curriculum development in modern distance education," in *Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pp. 1545–1548, Chengdu, China, December 2017.

[2] E. San Cristobal Ruiz, M. Tawfik, S. Martin et al., "Design, development and implementation of remote laboratories in distance electronics, control and computer subjects," in *Proceedings of the 2013 10th International Conference on Remote Engineering and Virtual Instrumentation (REV)*, pp. 1–5, Sydney, Australia, February 2013.

[3] M. Al-Adhaileh and A. Al Friday, "A flexible distance education delivery model: design and implementation at king faisal university," in *Proceedings of the 2015 Fifth International Conference on e-Learning (second)*, pp. 312–315, Manama, Bahrain, October 2015.

[4] L. Tobarra, S. Ros, R. Hernandez et al., "Low-cost remote laboratories for renewable energy in distance education," in *Proceedings of the 2014 11th International Conference on Remote Engineering and Virtual Instrumentation (REV)*, pp. 106–111, Porto, Portugal, February 2014.

[5] M. Chen and D. Tian, "The multimedia technology application in the network distance education - accountancy as a model," in *Proceedings of the 2015 7th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 673–677, Huangshan, China, November 2015.

[6] J. Wirbel, K. Zych, M. Essex et al., "SIAMCAT: user-friendly and versatile machine learning workflows for statistically rigorous microbiome analyses," *bioRxiv*, 2020.

[7] D. Hooshyar, M. Pedaste, K. Saks, Ä. Leijen, E. Bardone, and M. Wang, "Open learner models in supporting self-regulated learning in higher education: a systematic literature review," *Computers & Education*, vol. 154, Article ID 103878, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5374832, 11 pages
https://doi.org/10.1155/2021/5374832

Research Article

Construction and Analysis of Intelligent English Teaching Model Assisted by Personalized Virtual Corpus by Big Data Analysis

Jinxia Zhu,[1] **Changgui Zhu**,[2] **and Sang-Bing Tsai**[3]

[1]*Shanghai Communications Politechnic, Shanghai 200431, China*
[2]*Zhuhai College of Science and Technology, Guangzhou, Zhuhai 519041, China*
[3]*Regional Green Economy Development Research Center, School of Business, Wuyi University, Nanping, China*

Correspondence should be addressed to Jinxia Zhu; zhujinxia20031@aliyun.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 30 August 2021; Revised 17 September 2021; Accepted 29 September 2021; Published 15 December 2021

Academic Editor: Xianyong Li

At present, a new round of scientific and technological revolution and industrial transformation with information technology at its core are accelerating. At present, a new round of scientific and technological revolution and industrial transformation with information technology at its core is accelerating. The challenge of new economy and new industry has put forward new requirements for the training of talents in China. The challenges of new economy and new industry have put forward new requirements for the cultivation of engineering talents in China. Based on corpus, this study constructed a model of intelligent English teaching assisted by virtual corpus. The traditional teaching of college English reading is based on, around, and for texts. Using DDL model, teachers can break the limitation of textbooks. On the basis of analyzing the general idea of the text, they can search out massive real corpus related to the general idea of the text by searching the core words in the text, so as to provide extensive reading resources for students in the maximum range. At the same time, teachers can rely on the corpus to design different types of teaching activities, realize student-centered task-based, inquiry-based, and autonomous learning and cultivate students' critical thinking ability, practical ability, and cross-cultural communication ability. This model breaks the limitation of "classroom + textbook," realizes student-centered task-based, exploratory, and autonomous learning, trains interdisciplinary new engineering talents needed by emerging industries and new economy in the future, and promotes the sustainable development of English teaching. Corpus-data-driven college English teaching mode breaks the limitation of "classroom + textbook," changes the traditional college English teaching mode, and realizes student-centered task-oriented, inquiry-based, and autonomous learning.

1. Introduction

Under the influence of globalization, English has become one of the necessary means to communicate with the outside world. Learning English can not only help students to broaden their horizons but also improve their cultural awareness and cultivate their thinking quality, which plays an important role in the development of each student. The rapid progress of modern information technology also provides a faster and more convenient way for English teaching and research [1–3]. Corpus used to retrieve language forms has been promoted in various fields of linguistics because of its rapid and accurate advantages.

Data-Driven Learning, referred to as DDL, is an active learning method. It is a "student-centered" discovery learning method, in which students bring questions to the corpus stored with real language use examples to find answers [4]. In this way, students can have a deeper impression on knowledge. From a more realistic language environment, academics can use examples to find themselves learning the use, value, and significance of language, thus improving the effect of foreign language learning, helping students to learn reflection, and improve their learning initiative. A large number of research results show that data-driven learning is an ideal learning method in foreign language vocabulary learning [5, 6]. In addition, corpus linguistics also has a

positive impact on the study of linguistics itself. The application of corpus linguistics in linguistics branches, such as phonetics, morphology, syntax and pragmatics, can carry out language research activities at more levels. For example, in terms of phonetics, we can carry out research activities on the way of foreign language pronunciation, and pronunciation and intonation of spoken language corpus. In the aspect of morphology, we can carry out researches on the composition and classification of words. In the field of sentence science, we can conduct research on sentence form and structure, discourse, and stylistic analysis. In pragmatics, research activities such as the use of language in the actual communicative context can be carried out [7]. The development of these research activities will help foreign language teachers master modern language rules and point out the direction for students to learn foreign language knowledge. In addition, the learning of corpus linguistics can help students understand the correct expression meaning and practical use of some words, and with the help of advanced computer technology, it helps students to realize the meaning of corpus linguistics words, sentence patterns, and other forms of language expression. It can be seen that the application value of corpus linguistics in foreign language teaching is relatively high [8–10].

With the development and deepening of corpus research, language teaching and language research have obviously formed a unified trend [11]. Language teaching based on corpus adopts the "3I" teaching mode, which advocates that language teaching consisting of three stages: Illustration, Interaction, and Induction. Illustration refers to learners observing real language data. Interaction means that learners discuss and share findings in the corpus. Induction refers to the rules that learners develop for a language point, which can then be improved on by observing more data [12]. The "3I" teaching method believes that language knowledge exists in large-scale real texts, and language rules should be discovered and summarized through the investigation of real corpus, which is a bottom-up exploratory experience induction. According to data-driven learning, effective language learning itself is also a kind of language research, and the observation of index rows can promote learners to master the strategies of inductive learning, especially the strategies of distinguishing similarities and differences, inferring hypotheses, and verifying them [13]. Data-driven study argues that the vocabulary and grammar of binary opposition expand the language meaning units as much as possible, according to the language concept, the concept of the term is far greater than words, from the node word collocation, and classes to join, to semantic orientation and the semantic rhyme, from concrete to abstract vocabulary syntax level, and an extension to the pragmatic level. The results of cnKI retrieval were imported into CiteSpace software for data conversion and visual analysis. Keywords not directly related to this study were removed, such as "corpus," "vocational English," and "vocational English," and then the rest of the high-frequency keywords were combined and classified. The research on the application of corpus methods in higher vocational English education can be roughly divided into nine topics, as shown in Figure 1: interlanguage research and English teaching, lexicology, corpus creation, corpus linguistics, pragmatics and cognitive linguistics, corpus technology, textbook research, language testing, and computational linguistics (Figure 1).

Personalized virtual corpus is based on the fully mix traditional corpus technology advantage, through the cloud and big data network technology to overcome the disadvantage of traditional corpus, using the Internet platform of endless resources, or specify a set of corpus part of the generic corpus as a basis of source data corpus [8]. The construction process of this kind of corpus is simple and easy to use. The most important thing is that it can meet the various personalized needs of users. Therefore, it is widely used in language research and learning, which can not only improve the level of language teaching but also provide sufficient corpus and innovative research perspectives for subsequent language research. Corpus is constructed mainly with massive language samples, and it has the characteristics of authenticity and diversity of content. With the help of the language features of real language samples, corpus linguists can provide real and effective reference for foreign language learning and teaching. In the foreign language classroom, the introduction of corpus linguistics theory provides students with rich corpus resources and real language use environment. Students can use the help of corpus to find answers to their questions, thus improving their learning initiative and playing a positive role in improving the effect of foreign language teaching. With the help of the corpus, teachers can choose the important and difficult points of the classroom content reasonably, which is conducive to the reasonable arrangement of the teaching content and the smooth progress of the class. In view of the advantages of corpus linguistics in foreign language teaching, it is of great significance to actively explore the reasonable combination of foreign language teaching and corpus linguistics, which is conducive to the reform and development of foreign language education.

2. Related Work

Applied corpus teaching has been extensively studied. Paek and Kim [14] divide corpus functions into two categories: one is that corpus reflects social interaction function to some extent. The other is that corpus can improve the efficiency of language processing. Corpus as multifrequency words in our daily communication. If we can properly understand and use the corpus, it will help to improve our daily communication ability. So far, corpus teaching has highlighted its key role in foreign language teaching due to its *s* structure collocation, reasonable syntactic rules, and limited language environment, and it has injected new ideas into foreign language teaching. In daily social communication, it is found that corpus appears very frequently in oral communication, which has a certain influence on the composition of sentences. Gowin-Jones [15] pointed out the enlightenment of corpus structure on English teaching and proved that corpus teaching method is conducive to learners' foreign language acquisition through exploration. The corpus conforms to certain syntactic rules and can be stored in the brain as a

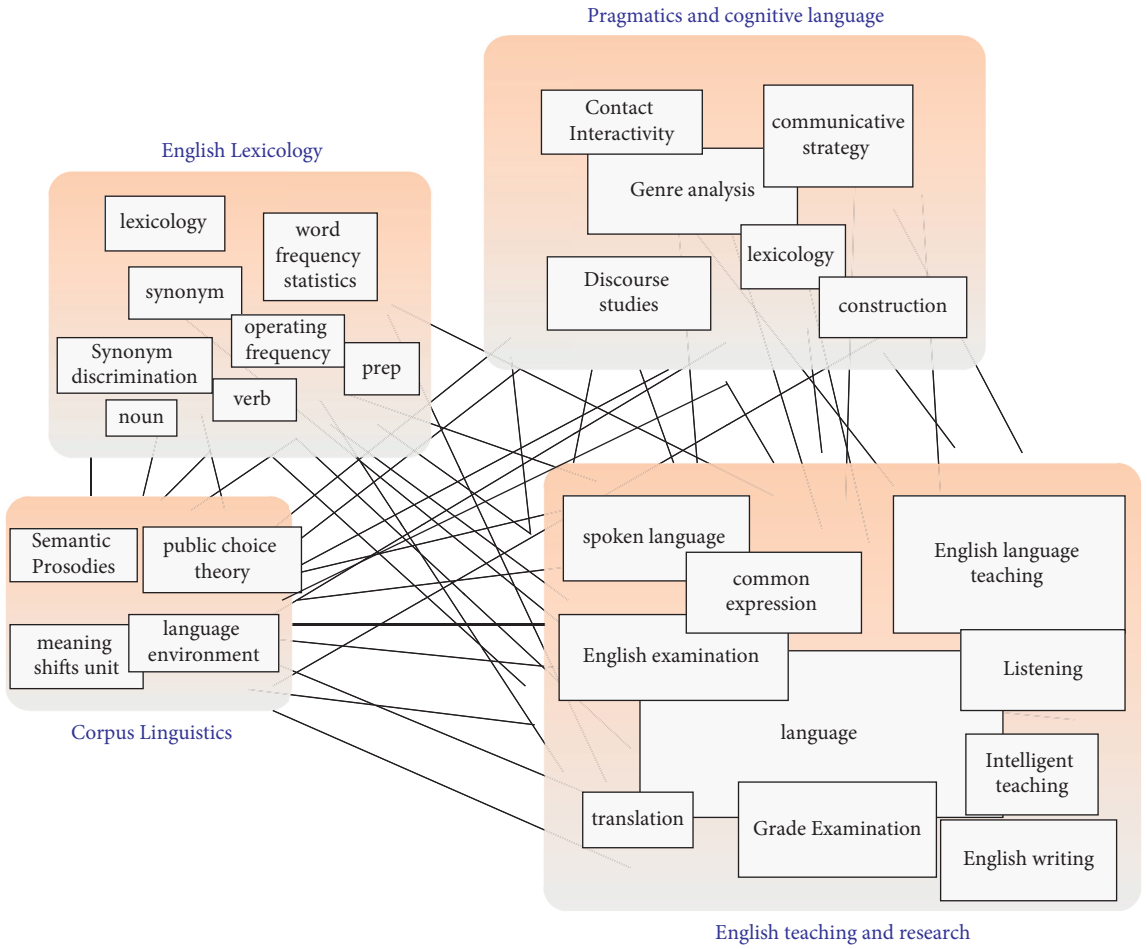


Figure 1: Topics of English corpus research.

whole. This not only relieves the communication pressure but also improves the expression ability of the communicator, which occupies an important position in oral English. However, in the process of English writing teaching, corpus method has not been paid much attention. The author believes that corpus teaching plays an active role in the writing process: corpus teaching helps to improve the accuracy and fluency of learners' use of corpus. Bibauw et al. [16] found that the input and output of high-frequency chunks are the key to improving learners' fluency in foreign language writing. It can be further proved that chunk teaching method is conducive to improving the language competence of foreign language learners in China. Therefore, language learners can use chunks to improve their writing skills. Huo [11] points out that every native speaker has a lot of linguistic blocks stored in his mind, so he can use the language more accurately and fluently in daily communication. If learners master a certain number of idiomatic chunks, they can ensure more idiomatic and fluent language output, which is conducive to improving learners' language ability. AlShuweihi et al. [17] explored the combination of task-based teaching method and chunk teaching method in teaching, and the statistical results show that this teaching mode is helpful to cultivate students' autonomous learning ability and improve learners' pragmatic ability to a certain extent. Adikari et al. [18] explored a new teaching model that combines mind mapping with chunk theory, and the experiment proved that this model can improve the effectiveness of students' English reading to some extent.

3. Intelligent English Teaching Platform Based on Personalized Virtual Corpus

3.1. Data Source. The research data of this study are mainly derived from two corpora: One is the Chinese English learners corpus (CLEC) of college Cet-4 and Cet-6 (ST3 and ST4) and the other is the LOCNESS corpus of American College Students' essays (brusR1, BRSUR2, BRSUR3, USARG). The total number of words in the two corpora is 416 476 (CLEC corpus) and 245 321 (LOCNESS corpus), respectively. In addition, this study chose the word "RATHER" as the target word in the case to explore a new model of college English vocabulary teaching. In English writing and conversation, the word "rather" is used in a variety of different parts of speech. It can not only be used as an adverb of degree to enhance the author's mood and expression but also can be used with "would, than, but" and other words to express the author's subjective will and preference or the difference in objective existence. We have collected the subtitles of 112 excellent English films and

documentaries to build a film and television corpus with a wide range of subjects, including both classic business card and some excellent films in recent years, in order to fully reflect the characteristics of contemporary English. The main parameters of the corpus are shown in Table 1. At the end of the experiment, 45 students in the experimental class were given questionnaires, 45 of which were effectively recovered. The questionnaire is divided into four dimensions, and the questions in each dimension. It completely matches the five choices. SPSS19.0 was used to analyze the data.

Table 1: Main parameters of the corpus.

Assessment items	Amount
Total words	29840
Tokens words	1239103
Token ratio	40.02
Characters	4024392

3.2. Design of Learning Environment Based on Constructivism. The instructional design based on constructivism attaches great importance to the design of learning environment, which is also an important feature that distinguishes it from the traditional teacher-centered instructional design. Tellez and Villela [19] believed that learning environment is a place where learners can use various tools and information resources and cooperate and support each other in the pursuit of learning goals and problem-solving activities. In this view, various tools and information resources used in learning are also included in the category of learning environment, which further enriched the concept of learning environment. He proposed a constructivist learning environment model, which consists of six parts, as shown in Figure 2:

To understand any problem, learners need to have some experience on the problem and be able to construct the corresponding mental model. However, for ordinary beginners, what they lack most is experience, which is very critical for them to solve problems. Therefore, it is very important for a constructivist learning environment to provide a series of relevant examples that learners can refer to. Related instances in constructivist learning environments can support learners' learning in two ways: First, related instances in constructivist learning environments can help students memorize. When people encounter a problem or situation for the first time, they will naturally look for similar cases in their memory that they have solved before, compare their previous experiences and lessons with the current problem, and if the goals or conditions match, they will apply the method of solving the problem in the previous case. Therefore, relevant cases help or supplement memory by providing learners with representations of experience that they do not have.

Secondly, relevant examples in constructivism learning environment can enhance learners' cognitive flexibility. As a branch of constructivism theory, cognitive flexibility theory believes that traditional teaching often simplifies the real background of complex problems and gives students a one-sided understanding of the problem. Therefore, the theory advocates providing a variety of representations and explanations about content to show the complexity of the knowledge domain itself, the connections between concepts and concepts of a certain point of view, and the connections within the concept, and it also advocates to convey multiple points of view on many issues with multiple and related instances. Therefore, in order to emphasize students' cognitive flexibility, relevant examples should provide a variety of perspectives and perspectives on the problem to be solved.

The application of corpus-based intelligent technology in English education is mainly shown in the following aspects (as shown in Figure 3). From the preclass preview, interactive exercises, teaching resources, and other aspects of learning path planning for learners. Push personalized learning resources to learners according to their preferences and needs; through adaptive learning, interactive, intelligent, and accurate teaching and learning can be realized. Intelligent lesson preparation can provide massive teaching resources. In the intelligent teaching link, the "smart classroom" of cross system and interactive teaching is realized to achieve the seamless connection between teaching and learning. In the intelligent work, the visual class work situation is generated automatically by the intelligent system. Relying on the intelligent teaching and research platform, the construction of online teaching and research community is conducive to the realization of online and offline linkage and the improvement of teachers' professional skills (Figure 3).

3.3. Construction of Corpus-Assisted English Teaching Model. Corpus-assisted English teaching has the following steps:

(1) Under the guidance of teachers, students choose research questions according to the course content and interest. At the same time, the instructor should teach basic corpus retrieval methods to lay a good foundation for the following courses and activities.

(2) Make a complete study plan through group discussion and teachers' suggestions. This plan will provide students with evidence to carry out their study activities and make them understand the general research process.

(3) Activity exploration: through the use of various reference books, computer networks, and other ways to collect and read literature, at the same time, questionnaire survey, interview, field investigation and research, and other means to collect data and use relevant statistical software for data analysis and analysis.

Discuss and come to a conclusion. In this link, many language problems encountered by students can be solved through corpus retrieval, which effectively alleviates the problem of lack of teacher guidance after class.

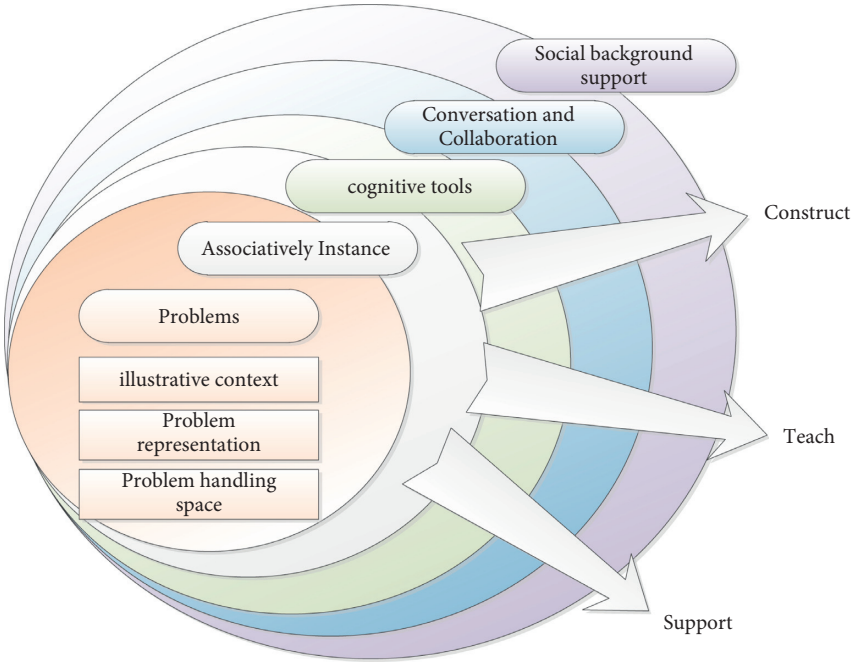


Figure 2: Constructivism-based learning environment model.

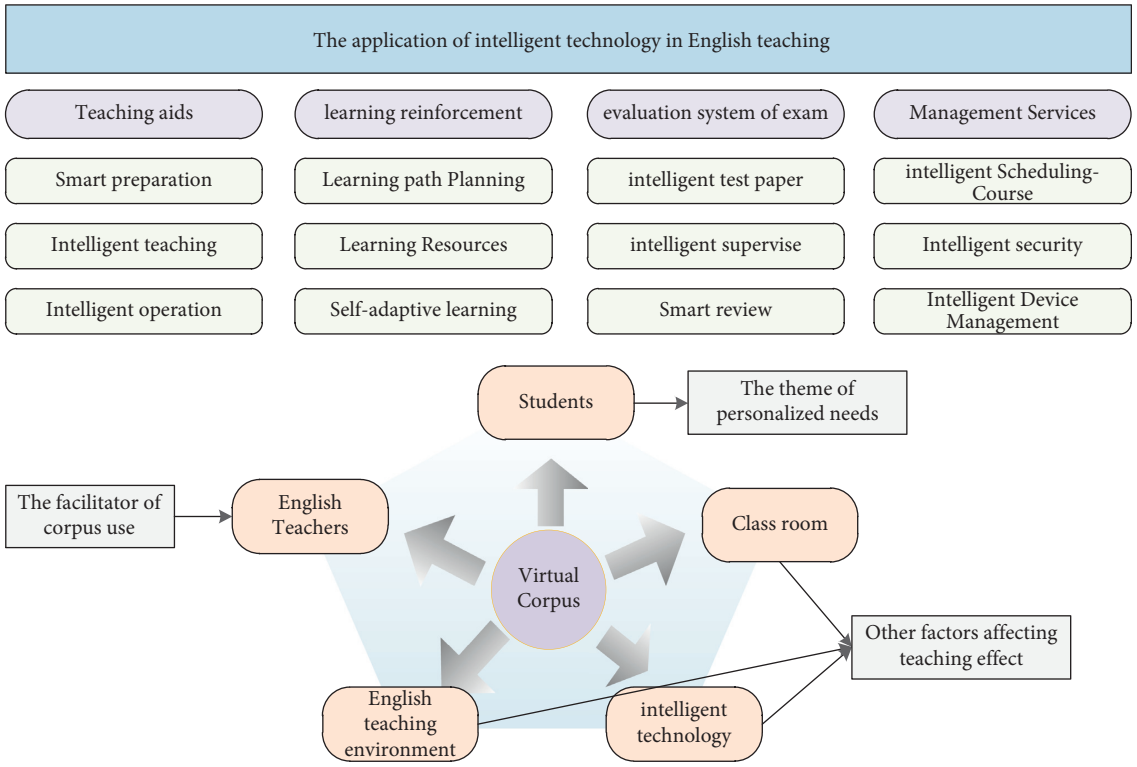


Figure 3: The relationship between corpus and other teaching elements.

(4) Listen to the guidance and suggestions of teachers and other students, find out the problems and deficiencies in the research in time, and make improvements in the follow-up research.

(5) Project report writing. Students summarize their research topic and submit it in the form of research report. This process can improve students' English academic writing ability and enhance their sense of

academic style. At this stage, the corpus plays a crucial role. A series of language problems, such as vocabulary, grammar, sentence pattern, and collocation, encountered by students in the process of writing can be solved through corpus retrieval, observation, analysis, and summary, especially 'students' writing style awareness can be greatly improved.

Traditional college English reading teaching is text-based, text-centered, and text-specific teaching activities. By adopting the intelligent teaching mode, teachers can break the limitation of textbooks, on the basis of analyzing the main idea of the text, search the massive real corpus related to the main idea of the text by searching the key words in the text, and provide extensive reading resources for students in the largest range. At the same time, teachers can design different types of teaching activities based on corpus to cultivate students' critical thinking ability, practical ability, and intercultural communication ability.

3.4. Construction of Personalized Virtual Corpus. The reason why personalized virtual corpus has become the most commonly used form of corpus at present is inseparable from its own outstanding advantages. First of all, virtual corpus can use Internet resources to build a small corpus related to a certain topic to meet the personalized needs of users, especially suitable for the study of ESP. Secondly, the virtual corpus is easy to use. It takes about 3-4 seconds to build a virtual corpus. In addition, there are corresponding construction guidelines on many platforms, and users can easily build their own personalized virtual corpus according to the guidelines [20–22].

First of all, a corpus platform website, http://corpus.byu.edu/wiki/, login to register. Second, click "Search" at the top left of the main page and click "Create Corpus" to enter the page of Corpus creation. Third, type in the subject word, which is the core word of the virtual corpus to be created, such as "Flight Procedure," and then single click "Find matchingstrings," and you will Find that the page displays everything related to that term. Next click SORT/LIMIT and then click "Relevance." At the same time, the "MINIMUM" option can also be used to screen out the corpus with the lowest occurrence frequency of keywords in the text, so as to facilitate further editing of the corpus. Finally, click "Save List" to name and save the corpus, thus completing the construction of a small virtual corpus. At the same time, the corpus can be further edited and deleted, and users can complete different language research purposes according to different needs.

Content-based recommendation algorithm mainly makes recommendations by obtaining the attribute characteristics of people or things. The advantage of this algorithm is that it can recommend items to users according to their preference characteristics, thus solving the recommendation problem of new items. Therefore, the selection of the similarity function is very important. By calculating the similarity between individuals, we can accurately find the n neighbors of the nearest neighbor. The similarity between users is calculated according to the similarity function, and then, a set of nearest neighbor users (excluding target users) is obtained by sorting the similarity degree from large to small. The calculation formula of Euclidean distance similarity algorithm [3, 4] is given as follows:

$$\omega_{ij} = \sqrt{\omega_1 \cdot \left(S_{i,1} - S_{j,1}\right)^2 + \omega_2 \cdot \left(S_{i,2} - S_{j,2}\right)^2 + \cdots + \omega_n \cdot \left(S_{i,n} - S_{j,n}\right)^2} + g_0. \quad (1)$$

Application scenarios of intelligent English teaching evaluation mainly use intelligent diagnosis evaluation system to analyze students' learning situation before class and activate students' existing knowledge and experience. Analysis of class and individual classroom practice data and timely feedback are remedy. After class, we can timely detect the learning effect through online homework to review and strengthen. Thus created algorithm includes listening, speaking, reading, and writing comprehensive intelligent English teaching evaluation application scenarios, let the students before class, during class, and after class do not deviate from the teaching content of learning at all levels of education, on the basis of understanding English teaching content background cultural reinforcement accurate pronunciation, understanding the linguistic logic of teaching content, and learn to control self-assessment, and truly help students build their own knowledge system from preview to review after study (as shown in Figure 3).

$$\omega_{ij} = \frac{\overrightarrow{S_{i,n}} \cdot \overrightarrow{S_{j,n}}}{\left\|\overrightarrow{S_{i,n}} - \overrightarrow{S_{j,n}}\right\|} = \sin\left(\overrightarrow{S_{i,n}}, \overrightarrow{S_{j,n}}\right). \quad (2)$$

The similarity between the vectors of user Si (n) and user Sj (n) is expressed [12] by the cosine value of the angle between the space vectors, and the similarity between the vectors is proportional to this value.

The calculation formula of Pearson similarity algorithm is given as

$$\omega_{ij} = \frac{\sum_k \left[S_{i(k)} + S_{j(1)}\right]\left[S_{i(k)} + S_{j(1)}\right]}{\sqrt{\left(S_{i,1} - S_{j,1}\right)^2 + \left(S_{i,2} - S_{j,2}\right)^2 + \cdots + \left(S_{i,n} - S_{j,n}\right)^2}}. \quad (3)$$

4. Result Analysis

4.1. Statistics of Basic Parameters of the Corpus. Concordance 3.2 contains some statistical functions for the basic parameters of the corpus. Figure 4 shows the statistical results of the basic attributes of the corpus of texts 1–6 in the Intensive reading textbook of College English by Using Concordance 3.2:

These include types, tokens, type-token ratio, words/sentences, and other parameters mentioned in the introduction of common statistical parameters of corpus. Word length distribution function can be used to calculate the word length distribution of corpus. This parameter has important reference value for judging the difficulty and language style of corpus text. Figure 4 shows the word length distribution of the corpus of texts 1–6 in the intensive reading textbook of College English (Figure 4).

It can be seen from the figure that there are 235,364 graphic signs with word length of 4, accounting for 18. The total number of graphic signs was 39,149 graphic signs with word length accounting for the total number of graphic signs.

4.2. Comparison between Intelligent English Grammar Teaching and Traditional Teaching. Through the analysis of the results of the pretest, it is found that the English grammar level of the two classes is similar before the experiment. This eliminates the interference factors for the experiment, as shown in Figure 5. By comparing the posttest scores of the two classes, it was found that students in the experimental class with flipped classroom teaching model had a greater improvement in English grammar (Figure 5).

The posttest results of the control class of the experimental class compared with the *t*-test results of independent samples showed that the t-distribution value of the test was 2.508, and the corresponding significance SIG value was 0.014, which was less than 0.05 to reach the significance level. Therefore, there was significant difference between the results of the posttest results of the control class of the experimental class. According to the statistical results of the classification mean, the posttest score of the experimental class was 38.53, which was significantly higher than that of the control class (35.07). It can be seen from this that compared with the traditional grammar teaching mode, flipped classroom teaching mode has made great progress in students' grammar scores.

Through the analysis of the test results of the control class, it is found that the students' grammar level has been improved under the traditional grammar teaching mode, but the effect is not obvious. Through the analysis of the test results of the experimental class, it is found that the

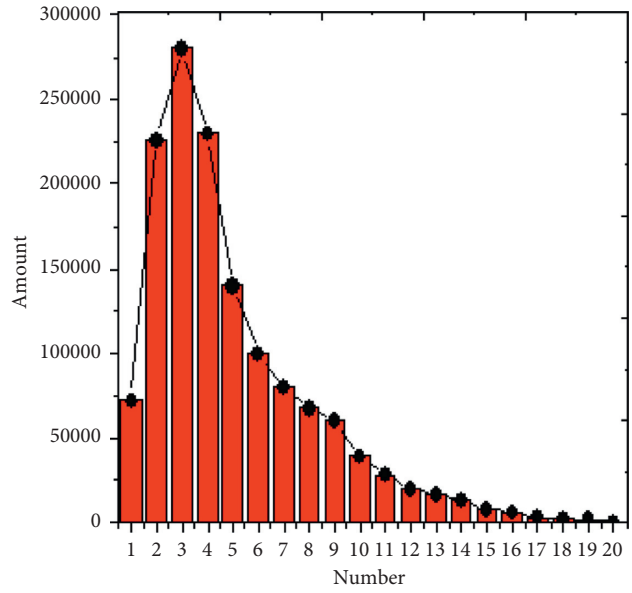


Figure 4: Statistics of basic parameters of the corpus.

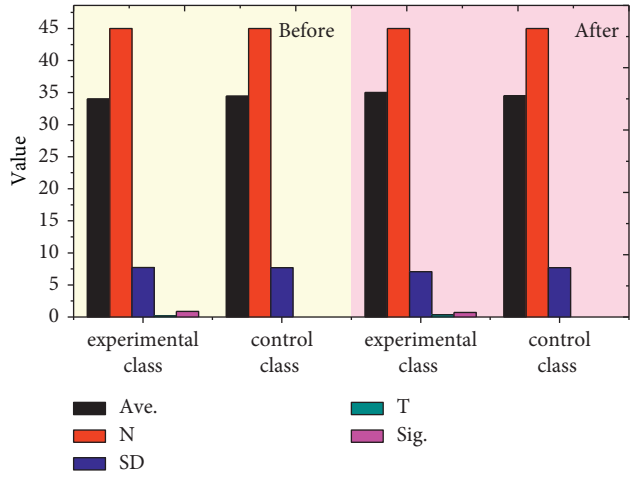


Figure 5: Comparison between intelligent English grammar teaching and traditional teaching.

application of flipped classroom teaching mode has greatly improved students' English grammar level, and the overall average score of the class has increased by 4.24 points. In conclusion, the flipped classroom teaching model based on ARCS model can improve students' grammar application.

Let the error cost function of student S_1 in question $K1$ be

$$J(S_1) = \theta^{10} - \theta^{11}\alpha_{11} - \theta^{12}\alpha_{12} - \cdots - \theta^{1n}\alpha_{1n} + g_{11}. \quad (4)$$

Similarly, the error cost function of student S_1 in question $K2$ is

$$J(S_1) = \theta^{20} - \theta^{21}\alpha_{21} - \theta^{22}\alpha_{22} - \cdots - \theta^{2n}\alpha_{2n} + g_{20}. \quad (5)$$

Therefore, the average error cost function of the predicted score and the real score of the S_x student is

$$J(S_x) = \theta^0 + \left[\sum_{m=1}^{m} \theta^0 \alpha_0 + \theta^n \alpha_n + \sum_{m}^{m} g_{m0}\right]. \quad (6)$$

4.3. Corpus-Based Analysis of English Teaching Scenarios. In this teaching mode, flipped classroom is divided into two parts: preparation stage and implementation stage. The preparation stage is mainly set for the teacher, including three main links: selecting the course before class, designing the syllabus plan, and recording the teaching video. The implementation stage of flipped classroom is mainly targeted at classroom teaching setting and goal oriented. Relevant knowledge is reviewed through the review of teaching videos, knowledge is mastered through tests and group activities, and knowledge is condensed and refined through summary.

Application scenarios of intelligent English teaching evaluation mainly use intelligent diagnosis evaluation system to analyze students' learning situation before class and activate students' existing knowledge and experience. Analysis of class and individual classroom practice data and timely feedback are remedy; After class, we can timely detect the learning effect through online homework to review and strengthen. Thus created algorithm includes listening, speaking, reading and writing comprehensive intelligent English teaching evaluation application scenarios, let the students before class, during class, and after class do not deviate from the teaching content of learning at all levels of education, on the basis of understanding English teaching content background cultural reinforcement accurate pronunciation, understanding the linguistic logic of teaching content, and learn to control self-assessment, and truly help students build their own knowledge system from preview to review after study (as shown in Figure 6.

The intelligent diagnostic evaluation system makes the diagnostic evaluation system change from the original single teacher experience evaluation, diagnosis and teaching management to data-supported teaching evaluation, diagnosis and teaching management, and changes from static intervention to dynamic and adaptive intervention, such as teachers can see the students' filling in the blanks after "Listen, Readandsay" in part A of the textbook. Compared with the previous teachers who only listened to most people's answers or asked questions, the intelligent diagnosis and evaluation system can immediately feedback the answers when the students answer the questions, so that the teachers can timely adjust the content of the teaching. Teach students more precisely where they are weak.

The evaluation process of intelligent diagnosis includes the whole process of pre-evaluation before test, post-evaluation after summary, and real-time data evaluation before, during, and after class [13].

4.4. Application and Analysis of Intelligent Teaching Platform. In order to test the practical effect of English wisdom classroom teaching, explore the shortcomings of this model, summarize experience, and lay an effective practical basis for the development of wisdom classroom teaching model in this school. We conducted a questionnaire survey. A total of 20 questions were set in the questionnaire, and the content of the survey also included multiple dimensions. The five-scale Likert method was adopted to design the questionnaire as follows: Strongly agree, agree, General, Disagree, and strongly Disagree. The single or multiple choice questions were mainly used, and the last question was an open question, as shown in Table 2.

According to the survey, about 75% of secondary vocational students like business English Wisdom classroom teaching mode, only a very few 4.86% disagree with it, 36.76% of students think that this mode solves the shortcomings of traditional classroom teaching, whereas 44.32% of students like to use the learning tools of this platform. Most students want to apply this model to other classroom activities. It can be concluded that the majority of students have a high evaluation of the model, fully showing the advantages of the model, and they also want to change the traditional model, integrate technology into the classroom, and stimulate personal interest (Figure 7).

The analysis in Figure 7 shows that, in terms of learning interest, 41.08% of the students think that learning activities in smart class stimulate their interest in learning business English, 12.43% of the students hold the opposite opinion, and only 4.86% of the students think that this activity cannot arouse their interest in learning this course at all. In terms of answering questions, 69.19% of the students were able to solve the problems they encountered in the learning process through the intelligent platform, whereas only 4.32% of the students disagreed. The comprehensive analysis of various factors and data shows that most secondary vocational students can accept this teaching mode, and a few do not agree with it, which may be because there are some obstacles in the learning process such as hardware and software problems, which lead to dissatisfaction with this mode. In terms of creating a relaxed learning environment in the classroom, about 70% of the students agreed with the relaxed and happy learning environment that smart classroom can create, whereas only 4.86% of the students said they dislike this kind of learning environment very much. After the wisdom of classroom teaching mode, the effect of the most important is whether the testing platform of learning resource wisdom deepen the understanding of knowledge; the data show the learner's point of relative balance: 14.59% of the learners do not agree with and about 60% of the learners think teachers classroom learning resources released by wisdom can further promote their understanding of knowledge. Therefore, in the future learning process, the release of teaching resources should be more in line with the practical needs of learners, and the content of resources should be more targeted.

The introduction of intelligent classroom teaching mode has greatly improved the participation of secondary vocational students in business English courses. As can be seen from Figure 8, nearly half of the students believe that the smart classroom platform can improve the frequency of interaction between teachers and students. It is understood that under the previous teaching mode, many secondary vocational students do not actively participate because of the

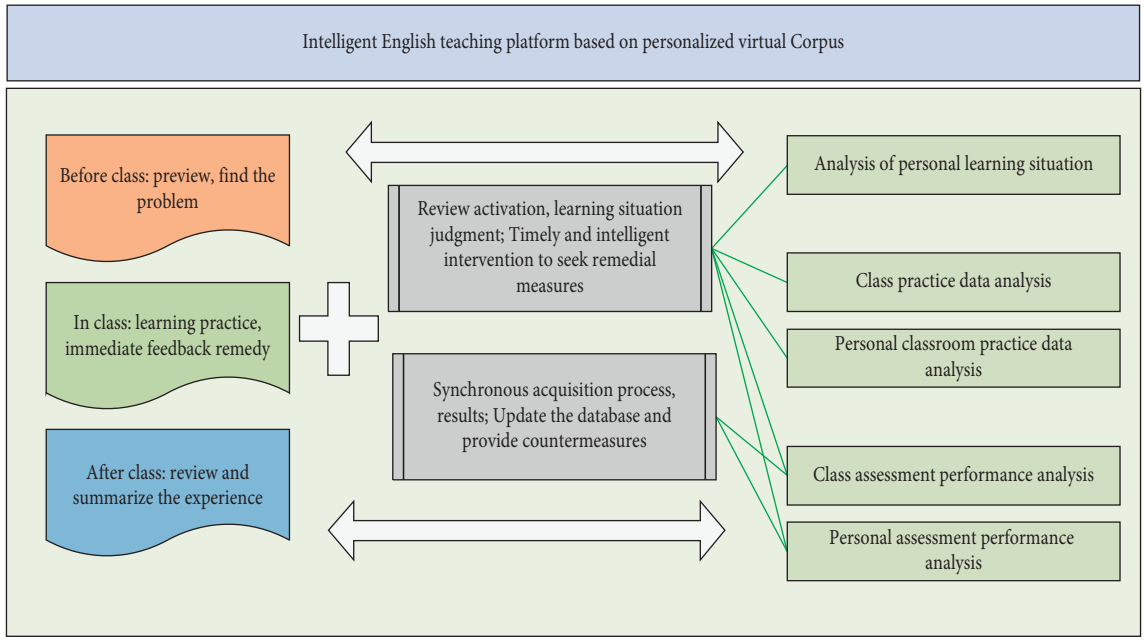


Figure 6: Corpus-based analysis of English teaching scenarios.

Table 2: Questionnaire survey index.

Dimensionality	Content dimension	The title number
Basic information	Age, class, etc	3
Learners' recognition of intelligent classroom teaching	4–7	4
Learners' satisfaction with classroom design	8–11	4
Learner engagement in the classroom	11–19	10
Views on intelligence corpus	20	1

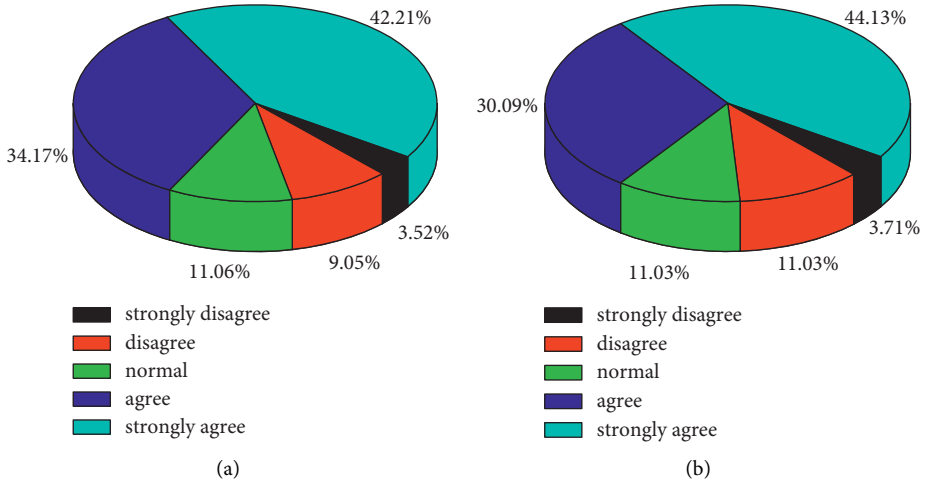


Figure 7: Continued.

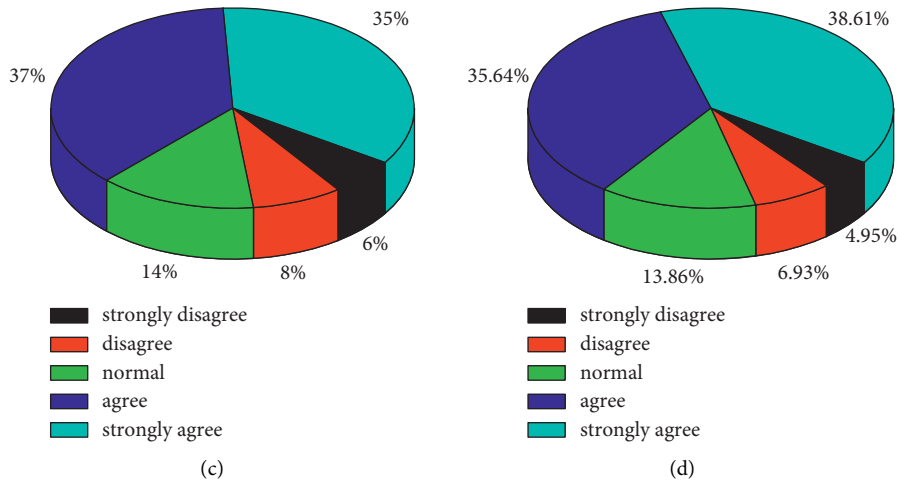


Figure 7: Academic recognition of English wisdom classroom teaching mode.

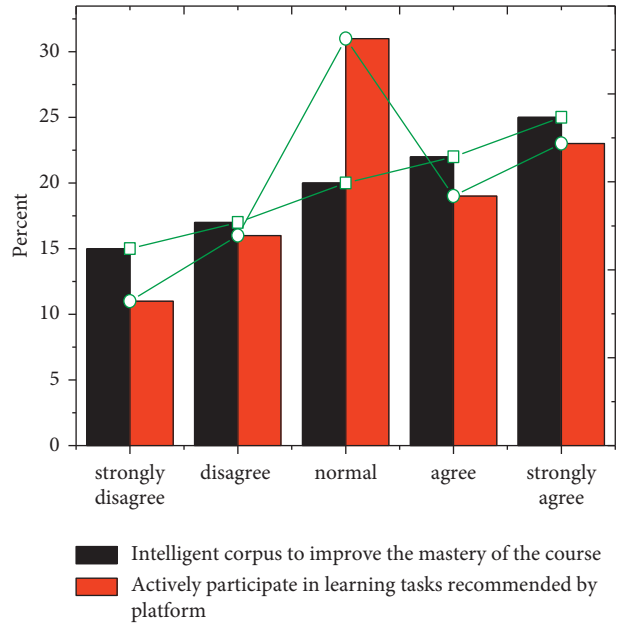


Figure 8: Student engagement in the virtual classroom.

class students, courage, and other reasons, resulting in less interaction with teachers, but through the wisdom platform, students can timely and effectively communicate with teachers; 41.62% of the students actively completed the learning task pushed by the teacher on the platform, 30.81% of the students generally agree with it, and only a few students do not agree. It can be seen that the smart platform effectively improves students' participation in class and their interest in learning (Figure 8).

5. Conclusion

In the age of intelligence, according to the core quality of English subject, artificial intelligence technology has supported every link of English education. This study established an English teaching platform based on personalized virtual intelligence. The process of applying artificial intelligence technology to English teaching effectively from the aspects of English listening, speaking, reading, and writing assisted by the platform to teaching, teacher research, and other aspects, makes English teaching more effective in cultivating students' core competence in English subject and cultivating students' English cultural awareness. Corpus-data-driven college English teaching mode broke the limitation of "class + textbook," changed the traditional mode of college English teaching, has realized the student-centered task type, exploratory, autonomous learning, is to explore to adapt to the economic development of new technologies, new industries, and new engineering education system with Chinese characteristics

of beneficial attempt. Through teaching practice, it is found that more attention should be paid to the following three problems in future teaching: (1) Free corpora such as COCA and BNC are monolingual corpora, which are difficult for learners, especially non-English majors; ② Students often get bored and confused by the complicated original search results. Therefore, we should pay full attention to students' interests and give them more encouragement. (3) Using DDL mode, teachers need to carefully design teaching activities and exercises, create teaching micro text, the workload of teachers is very large, so its large-scale promotion needs to have a strong teaching team support, to carry out cooperative lesson preparation.

In the future, we will carry out deeper and more extensive DDL college English teaching practice, and on the basis of empirical research, in quantitative research methods to explore the effect of the model to improve the students' comprehensive quality, constantly summarize and interactive teaching mode guided by teachers, in order to help Chinese learners to open up a new, highly effective, the wisdom of English learning.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This paper was supported by the Research on the English Translation of Chinese Idioms (Shanghai Jiaotong Vocational and Technical College 2018-2019 Academic Year Faculty-level Research; project no. JY1820); Reform and Practice of English Teaching in the Context of Curriculum Civics (Shanghai Jiaotong Vocational and Technical College 2019-2020 Academic Year Faculty-level Research; project no. JYZ2001).

References

[1] M. Sun and Y. Li, "Eco-environment construction of English teaching using artificial intelligence under big data environment," *IEEE Access*, vol. 8, pp. 193955–193965, 2020.

[2] X. Chen, D. Zou, H. Xie, and G. Cheng, "Twenty years of personalized language learning," *Educational Technology & Society*, vol. 24, no. 1, pp. 205–222, 2021.

[3] Y. Bin and D. Mandal, "English teaching practice based on artificial intelligence technology," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 3, pp. 3381–3391, 2019.

[4] A. Latham, K. Crockett, and D. McLean, "An adaptation algorithm for an intelligent natural language tutoring system," *Computers & Education*, vol. 71, pp. 97–110, 2014.

[5] X. Lu and B. Chen, *Computational and Corpus Approaches to Chinese Language Learning: An Introduction*, Springer, Singapore, 2019.

[6] T. Heift and M. Schulze, "Tutorial computer-assisted language learning," *Language Teaching*, vol. 48, no. 4, pp. 471–490, 2015.

[7] Z. Dong, "On the risk economic crime and its identification," *International Core Journal of Engineering*, vol. 6, no. 2, pp. 49–54, 2020.

[8] S. Y. Chen and J. H. Wang, "Individual differences and personalized learning: a review and appraisal," *Universal Access in the Information Society*, vol. 48, no. 1, pp. 1–17, 2020.

[9] B. Zhang, "Construction and application of the english corpus based on the statistical language model," in *Proceedings of the International Conference on Frontier Computing*, pp. 665–670, Springer, Singapore, 2018.

[10] T. Kabudi, I. Pappas, and D. H. Olsen, "AI-enabled adaptive learning systems: a systematic mapping of the literature," *Computers & Education: Artificial Intelligence*, vol. 2, Article ID 100017, 2021.

[11] Y. Huo, "Analysis of intelligent evaluation algorithm based on English diagnostic system," *Cluster Computing*, vol. 22, no. 6, pp. 13821–13826, 2019.

[12] X. Chen, D. Zou, H. Xie, and F. L. Wang, "Past, present, and future of smart learning: a topic-based bibliometric analysis," *International Journal of Educational Technology in Higher Education*, vol. 18, no. 2, pp. 1–29, 2021.

[13] L. Mora, R. Bolici, and M. Deakin, "The first two decades of smart-city research: a bibliometric analysis," *Journal of Urban Technology*, vol. 24, no. 1, pp. 3–27, 2017.

[14] S. Paek and N. Kim, "Analysis of worldwide research trends on the impact of artificial intelligence in education," *Sustainability*, vol. 13, no. 14, p. 7941, 2021.

[15] R. Godwin-Jones, "Data-informed language learning," *Language, Learning and Technology*, vol. 21, no. 3, pp. 9–27, 2017.

[16] S. Bibauw, T. François, and P. Desmet, "Discussing with a computer to practice a foreign language: research synthesis and conceptual framework of dialogue-based CALL," *Computer Assisted Language Learning*, vol. 32, no. 8, pp. 827–877, 2019.

[17] M. AlShuweihi, S. A. Salloum, and K. Shaalan, "Biomedical corpora and natural language processing on clinical text in languages other than English: a systematic review," *Studies in Systems, Decision and Control*, vol. 10, no. 1, pp. 491–509, 2021.

[18] A. Adikari, G. Gamage, D. De Silva, N. Mills, S.-M. J. Wong, and D. Alahakoon, "A self structuring artificial intelligence framework for deep emotions modeling and analysis on the social web," *Future Generation Computer Systems*, vol. 116, pp. 302–315, 2021.

[19] N. R. Téllez and P. R. Villela, "A personalized brand proposal based on user's satisfaction and curriculum supported by an intelligent job recommender system," *Radical Solutions for Digital Transformation in Latin American Universities: Artificial Intelligence and Technology 4.0 in Higher Education*, p. 217, Springer, Berlin, Germany, 2021.

[20] C.-C. Huang, "User's segmentation on continued knowledge management system use in the public sector," *Journal of Organizational and End User Computing*, vol. 32, no. 1, pp. 19–40, 2020.

[21] N. Ramu, V. Pandi, J. D. Lazarus, and S. Radhakrishnan, "A novel trust model for secure group communication in distributed computing," *Journal of Organizational and End User Computing*, vol. 32, no. 3, pp. 1–14, 2020.

[22] A. Aderonke, Oni, U. Musa, and S. Oni, "E-Revenue adoption in state internal revenue service: interrogating the institutional factors," *Journal of Organizational and End User Computing*, vol. 32, no. 1, pp. 41–61, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5341523, 11 pages
https://doi.org/10.1155/2021/5341523

Research Article

Research on the Optimization Method of Visual Effect of Outdoor Interactive Advertising Assisted by New Media Technology and Big Data Analysis

Zhinan Gan[1] **and Sang-Bing Tsai**[2]

[1]*School of Visual Arts, Hunan Mass Media Vocational and Technical College, Changsha, Hunan 410100, China*
[2]*Regional Green Economy Development Research Center, School of Business, WUYI University, Nanping, China*

Correspondence should be addressed to Zhinan Gan; gzn18692232209@163.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 6 September 2021; Revised 22 September 2021; Accepted 29 September 2021; Published 15 December 2021

Academic Editor: Xianyong Li

Along with the development and promotion of Internet technology, new media are increasingly diversified, enriching and changing our lives. This paper focuses on outdoor interactive advertising as a communication method, analyzing its creative features and applications. Unlike traditional advertising, where audiences receive passively, new media interactive advertising establishes a more direct communication method for consumers and products, creating different sensory stimulation and emotional experience environments according to the characteristics of products and conducting a series of interactions with consumers' sensory systems such as vision, hearing, smell, taste, and touch. The advertising method has also changed from passive reception to active participation. Finally, combining the characteristics of accurate delivery and the diversity of presentation forms of interactive advertising under the big data platform, the theme conception and form design of interactive advertising is proposed. The interactive design principles of highlighting the simplicity of the theme are explored. The interactive design principle of highlighting the simplicity of the theme is proposed, and the combination of color, text, and graphics is explored by the aesthetic visual design of the audience and the premise of focusing on the audience's feelings. Starting from the visual effect of outdoor interactive advertising, the change in the way advertising is conveyed is interpreting to us the arrival of a new communication concept, the arrival of the Internet-centered and technology-centered data integration era. The research in this paper aims to provide useful support for the diversified development of outdoor interactive advertising in the new media environment by exploring how to skillfully realize the creative expression of outdoor interactive advertising in the new media environment.

1. Introduction

In the new media era, advertising has ushered in a huge development opportunity, and the market position of new media advertising is rapidly rising. With the help of digitalization, informatization, and virtualization technologies, new media advertising has broken through time and space limitations, bringing users a new media experience and feeling, greatly improving audience brand recognition, and better-achieving marketing and promotion goals [1]. Advertisers' attitude toward new media has also changed profoundly, gradually shifting from passive acceptance to active use, looking for the best channel to showcase themselves on the Internet platform and reach a wider target audience group. It can be said that the rise of new media advertising has a disruptive impact on both advertisers and the market. For advertisers, as the market becomes increasingly competitive, it is necessary to use advertising to promote products to achieve sales revenue goals. Traditional advertising media are geared toward all audiences, without segmentation and classification, making advertising less effective [2]. With new media, advertisers can segment their audiences and establish communication and interaction mechanisms with them. For example, if a screen display is installed on a shopping cart when a customer goes to a certain product area, the screen will broadcast the

corresponding product advertisement, providing consumers with product introduction and reference information, which can guide their purchasing behavior, on the one hand, and help businesses collect data on customer preferences on the other [3]. Adopting this precise advertising information delivery method can effectively explore the potential needs of users and obtain the best communication effect with the lowest cost. In the industry, the renewal of media forms is the inevitable result of the development of social civilization, and people always hope to find an effective way to optimize the allocation of resources [4].

Based on the horizontal and vertical cross-border integration to form a new research field, advertising innovation in the new media context includes both the creative reconstruction of new media in terms of technology and concepts. The inquiry of advertising innovation is not limited to the field of creation, but a comprehensive innovation of the whole communication method, creation mechanism, and consumer interaction behavior [5]. Combining theory and practice, this paper uses typical outdoor interactive advertising cases to summarize detailed characteristics and specific analysis from two aspects: five-sense interaction and emotional experience, providing vivid materials and application attempts for interactive innovation theory research, which has a positive effect on the space for new media interactive advertising to carry out innovation and also has certain guiding significance for the creation and dissemination of digital art. This paper studies the survival value of outdoor advertising as well as new media through the social background and the survival status of outdoor advertising and explores the unique communication value, market value, and aesthetic value of outdoor advertising, while the integration of new media and outdoor advertising brings new media value to it. No matter which advertising media or which form of advertising is used, the purpose of advertisers is to spread the message and get feedback from the audience. It is hoped that the study will provide some theoretical support and reference for advertisers and advertising designers [6].

This paper is mainly based on the existing research of new media advertising, to explore the current situation of new media advertising design and then put forward new ideas of visual dynamic design of new media advertising. When creating interactive creative outdoor advertising design, adhering to the core of human-centered design concept, try to consider using one or more of the above-mentioned design techniques in the design to achieve a better publicity effect. Interactive outdoor advertising itself is designed to attract attention and draw potential customers closer to the communicator. Skillful mastery and use of the above design theories and methods will help designers grasp interactive advertising, improve the level of design, and solve the obstacles in design thinking. The first chapter is an introduction. This part firstly introduces the background and significance of outdoor interactive advertising with new media technology, points out the main methods of outdoor interactive advertising at present, and finally summarizes the research content and focus of this paper and lists the organization of the paper. Chapter two is related work. The chapter provides a reference for the selection and improvement of the research of this paper through the analysis of the current research status of outdoor interactive advertising under the new media. Chapter three is the research on the optimization method of outdoor interactive advertising visual effect based on new media technology. Through the research on the design method of outdoor interactive advertising under new media, the optimization model of outdoor interactive advertising visual effect is constructed, and then the outdoor interactive advertising design is evaluated. Chapter four is the result analysis. Through the analysis of optimization methods, quantitative evaluation analysis, and optimization effect analysis, the effectiveness of the methods studied in this paper is demonstrated, which can reflect the effectiveness of the optimization methods of outdoor interactive advertising visual effects based on new media technology. Chapter five is the conclusion. This chapter gives a general summary of all the contents of the article and gives a clear explanation of the optimization effect of outdoor interactive advertising, points out the main achievements of the article's research, and also points out the existing shortcomings according to the problems in the research process and provides an outlook on the future development direction of outdoor interactive advertising.

2. Related work

The research on new media advertising is also at the stage of active exploration. Although there are many excellent cases of new media advertising design, there is less research on the visual dynamic design theory of new media advertising, and the research results mainly focus on the meaning, definition, characteristics, and development of new media [7]. Ozcan and Hannah put forward the trichotomy of old media, new media, and new media. Old media is the media before the Internet was created, and the main communication characteristic is top-down control [8]. Stankov U et al. emphasize the establishment of a new relationship between brands and audiences through advertising that conveys product messages and is liked by the advertising audience [9]. The contextual advertising mentioned in the book involves outdoor media forms, citing a large number of outdoor advertising cases, using the spatial characteristics of outdoor media coupled with unique advertising creativity to make advertising a public space art to enhance the interactive experience with consumers, which plays an important role in enhancing brand image [10]. In the study of interactive advertising, if interactive advertising appears as a single form of advertising model, it often faces the problems of the small audience, inability to spread publicity, and low-value conversion due to the limitation of the base; then, in response to these problems, it is precisely ordinary advertising that can fill them. Together, they can reduce the unnecessary cost waste caused by their respective drawbacks and increase the interactive function of the advertisement and consumers based on the visual ornamental nature of the ordinary advertisement, to play the role of optimizing the actual effect of advertisement [9]. The update of science and technology undoubtedly provides strong information feedback and

technical support for the innovation of outdoor interactive advertising. With the continuous advancement of science and technology, interactive advertising will also integrate more communicative interactive methods.

The interactivity of outdoor advertising design is mainly to make the audience participate in outdoor advertising and meet their psychological needs. In modern times, as people receive and understand more information, audiences have long ceased to be satisfied with the passive installation of product information, and audiences are eager to turn the tables and dominate the dissemination of products [11]. Feng et al. derived the organic relationship between culture, human nature, urban architecture, and environment and outdoor advertising through their research and concluded that outdoor advertising has a close interaction with humanistic subjective factors such as culture, creativity, human nature, environment, and urban style personality [12]. Liu applied the conditional value assessment method and constructed a model of the relevant assessment system. The value assessment of open public space was studied in-depth [13]. Xie et al. conducted a detailed discussion on the value of outdoor advertising media, with a comparative study of the value of outdoor advertising media and commercial real estate [14]. The income approach in real estate was borrowed, and a comparative study was conducted, and a relevant model was constructed. Communication between outdoor advertising and audiences is particularly important in the new era, which is full of interactivity. Interactive advertising design is undoubtedly more eye-catching than traditional outdoor advertising design and is more likely to evoke consumers' desire to consume. Outdoor advertising design in the new era is given the ability to interact and be interactive by the new era of technology [15]. The new word represents the sublimation of science and technology and media materials, which is the transformation of the audience from the traditional single passive acceptance to the interaction between the audience and the communicator under the platform media based on the power of technology [16].

Through sorting and summarizing the relevant research literature and collecting cases of new media advertising, it is found that the current research on new media mainly focuses on the innovation of new media-related theories and the construction of theoretical systems, and the research on new media advertising mainly focuses on its types, characteristics, communication, development trends and design applications on the Internet, cell phones, outdoor new media, and other platforms, and there is a lack of relevant research and cases for new media advertising. There is a lack of research and case studies on the visual dynamic design of new media advertising [17]. The relevant research is also mainly focused on the connotation, definition, characteristics, and technology of new media advertising and is more successful in the practical application of new media advertising, with many excellent visual dynamic advertisements, but the relevant theoretical research is still relatively weak [18]. The research of experts and scholars in new media advertising has certain significance for this topic, and the clarified new media types, characteristics and advantages, the origin, development, and current situation of new media advertising, and the types and forms of new media advertising provide the certain theoretical basis for the research of the topic. Some typical cases of new media advertising provide a certain realistic basis for the research of the topic [19].

3. Research on the Optimization Method of Outdoor Interactive Advertising Visual Effect Based on New Media Technology

3.1. Outdoor Interactive Advertising Design Methods. In outdoor print advertising design, visual images play an important role, as the scenery is seen by human eyes intuitively. People can associate different things through intuitive images to meet their brain imagination and inner needs. Outdoor advertising itself is a form of advertising with a strong sense of space, which needs to instantly attract the attention of outdoor space audiences, which requires outdoor advertising to be designed with a larger area and strong visual impact to enhance the expressive power of outdoor advertising and make up for its shortcomings. Visual perception is the main form of traditional outdoor advertising application, which usually uses the arrangement and combination of color, text, graphics, and other elements to create visually impactful advertising images. This visual impact refers not to the bloody and violent scenes that bring strong sensory stimulation, but to the creative images that leave a deep impression on the audience through humorous and interesting images that can cause them to resonate and think inside [20, 21].

The visual effect is the core element that affects the first impression of the audience, and the audience has visual fatigue for the flat, static visual images. Innovation of visual effects can give audiences a bright feeling the first time [22, 23]. The visual design of outdoor advertising in the scene era is breaking through = dimensional space and developing in the direction of multidimensionality and dynamism. New media technology has provided greater room for the design of outdoor advertising. The relatively traditional outdoor advertising in the past has changed its relatively single expression form and been given changes in sound, color, and shape. Under the scene property generation, the design of advertising images has realized a change from static to dynamic. The design process of outdoor interactive advertising is shown in Figure 1. It is necessary to consider the size of the advertisement and the proportion of the pattern, the integration of the background color of the advertisement and the promotion, and the light and dark requirements of the advertisement installation location for the pattern. Design satisfactory publicity advertisements for customers. The durability of the advertising material, the firmness of the installation, and the material used in the picture carrier in consideration of the cost are to be perfect [24].

In the context of the new media era, the Internet and various advanced technological devices have enhanced the

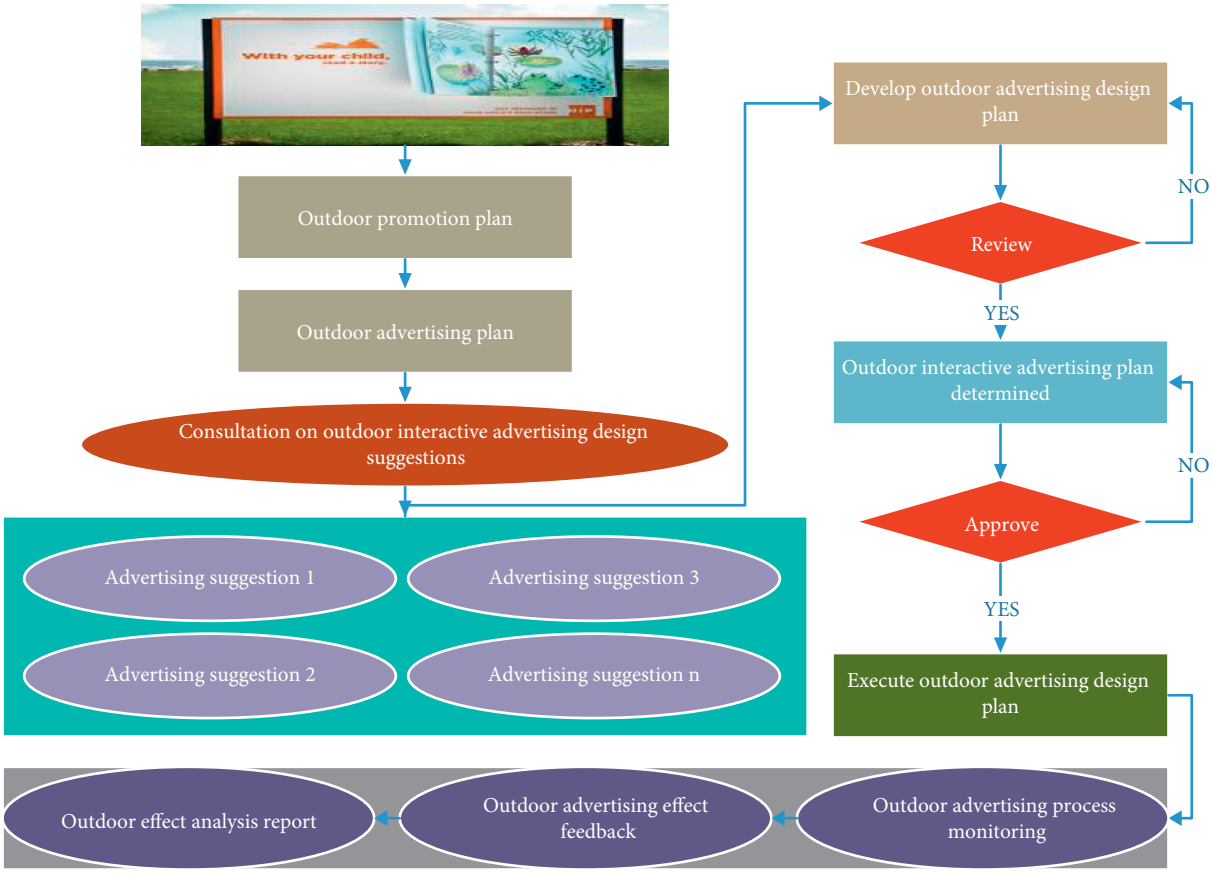


Figure 1: The outdoor interactive advertising design process.

connection between outdoor media and audiences and strengthened the interactive experience between them. First, the audience's initiative has increased. The influence of the Internet on audiences is greater than we can imagine. The openness brought by the Internet gives audiences the initiative to release information, and audiences are no longer willing to passively receive information, but they more often want to actively participate in the communication process. The emergence of new media technology, its advantages in information transmission and information communication, has completed the effective communication between advertisers and consumers, consumers and consumers, and consumers and advertising carriers, forming an interactive mode between them. From the perspective of communication science, interaction is the various behaviors of the transmitter and the recipient around information exchange, including sending and feedback. In the original communication mode of outdoor advertising, the advertiser has the absolute right to speak, and the audience only passively accepts the advertising information. It is difficult for the audience to give timely and effective feedback on the content of the advertising information. The combination of media technology has changed the one-way communication mode of advertising information. Consumers become the participants of advertising information, improving the status of the audience in the process of information dissemination, increasing the interaction between the information receiver and the publisher, and satisfying the audience's personalized demand for advertising information.

3.2. Visual Effect Optimization Model Design. The convenience and timeliness of new media have put forward higher requirements for the release of interactive advertisements, and it is often necessary to seize time and opportunities with many counterparts to leverage publicity. Advertisers need to make good use of and grasp the timing of advertising dissemination. Hue is the most easily distinguished and perceived color characteristic; saturation is related to the degree of evocation of color perception by users; and brightness reflects the overall tone of an image. To calculate the basic color characteristics, the average hue a, average saturation b, and average luminance c of all pixels in an image can be defined as shown in equation (1), where M and N are the width and height of the image, and $f(m, n)$ is the specific value of pixel point (m, n) under each color channel (hue/lightness/saturation).

$$G(a, b, c) = \frac{\sum_m^M \sum_n^N f(m, n)}{M * N}. \quad (1)$$

To define the color complexity in the image, this paper optimizes the quantization method for the color complexity measure. The color complexity of each pixel point (m, n) is defined as equation (2). $ß(m, n)$ is a custom sliding window,

(m, n) is the centroid of this window, and $f(m, n)$ is the average of the corresponding colors of this window and $f'(m, n)$. The difference between the colors is measured and normalized by the Gaussian function H.

$$\begin{cases} g(m,n) = \sum H\left(f(x,y) - f'(m,n)\right), \\ (x,y) \subseteq \beta(m,n). \end{cases} \tag{2}$$

The color difference is calculated as in equation (3), where ζ is the normalization factor, and $G(f(x, y), f(m, n))$ is the Euclidean distance between two points in the HSV color space.

$$f(x,y) - f'(m,n) = 1 - \ln\left(\frac{G(f(x,y), f(m,n))}{\zeta}\right). \tag{3}$$

The local color of each print advertising image consists of various types of elements such as background, copy, and subject. Among them, the background elements often occupy the largest color area and have a very important influence on the overall color performance. Therefore, for a background element b, the average color hue i average saturation j and average luminance k can be taken as the basic local color characteristics, as shown in formula (4), where $G(b)$ is the number of pixels in the background element G. $f(m, n)$ is the value of the pixel point $(m, n$ in the background element corresponding to the color channel (hue/lightness/saturation)).

$$G(i,j,k) = \frac{\sum_{(m,n)}^{G(b)} f(m,n)}{G(b)}. \tag{4}$$

In the context of the new media era, the Internet and various advanced technological devices have enhanced the connection between outdoor media and audiences and strengthened the interactive experience between them. First of all, the audience's initiative has increased. The influence of the Internet on audiences is greater than we can imagine. The openness brought by the Internet gives audiences the initiative to release information, and audiences are no longer willing to passively receive information, but they more often want to actively participate in the communication process. The emergence of new media technology, its advantages in information delivery and information communication, has completed the effective communication between advertisers and consumers, consumers and consumers, and consumers and advertising carriers, forming an interactive mode between them. To make the pairwise comparison results cover all the data as much as possible under the premise of ensuring certain labeling quality, this paper proposes a pairwise comparison method based on grid selection, which needs to cover ω of all comparison pairs, as shown in equation (5), where m is all the comparison objects, $m(m-1)(m-2)/3$ is all the comparison times, x is the number of comparisons, ϖ is the candidate items per unit grid, and z is the number of objects from which the user needs to select.

$$f(x, y-z) = \frac{m(m-1)(m-2)}{3} * \omega. \tag{5}$$

After getting all the comparison results, it is necessary to quantify the style match of each image on this basis. For example, if a print advertisement image has a "romantic" style match of 65%, there is a 65% probability that this image will be labeled as "romantic" by the annotator compared to the other images in the dataset. In this paper, we estimate the style probability of an image based on the principle of great likelihood estimation. Specifically, the result of the pairwise comparison of two images can be described as equation (6), where p denotes the style category of the image, $f(m)$ and $f(n)$ denote the two images compared, and q denotes the comparison result of the two images.

$$G = \|p, f(m), f(n), q\|. \tag{6}$$

From project value analysis to target audience positioning, from consumer group definition to target audience psychological needs, the target positioning of new media interactive advertising should strive for precision, combining the psychological characteristics of the target consumer group and the characteristics of new media interactive communication, supplemented by strategic planning, material performance, and reasonable media communication mix, to achieve the maximum market development. Only then can the advertising communication achieve good economic and social effects.

3.3. Evaluation of Outdoor Interactive Advertising Design. Since new media interactive advertising is built on a variety of advanced network and media technologies, it also shows advantages in its effect evaluation that traditional advertising can hardly match. The interactive nature of new media allows target audiences to submit personal feedback directly, and advertisers can receive the feedback information and conduct effect evaluation within a short period; The new media effect evaluation relies heavily on technical means, which consumes relatively less human and material resources, and the corresponding advertising cost is also lower.

The communication process of advertisement is nothing but four stages: exposure to an advertisement, acceptance of information, change of attitude, and action. Accordingly, advertisers can evaluate the communication effect according to different advertising purposes and corresponding formulas. In the exposure stage, the audience coverage of the media is mainly evaluated, and the evaluation index is mainly the number of exposure times of the advertisement; in the information acceptance stage, the reach of the advertisement is mainly evaluated, and the evaluation index is mainly the number of clicks and clicks rate of the advertisement; in the attitude change stage, the psychological change of the audience is mainly evaluated, and the attitude toward the product and brand is mainly changed; in the action stage, the influence of the advertisement on the purchase decision of the target audience is mainly evaluated. The latter two stages mainly evaluate the number of conversions and conversion rates. The correspondence between evaluation indexes and formulas is shown in Table 1.

When outdoor advertising is placed, the relative visual retention time is different. What we mean by more eye-catching is a situation where the visual grade is relatively high. The visual grade mainly includes the following factors.

Table 1: Correspondence of assessment indicators and formulas.

Index number	Evaluation stage	Evaluation formula	Evaluation index
1	Contact advertising	Notice	Ad impressions
2	Receive the info	Interest	Clicks and click-through rate
3	Change attitude	Desire	Conversions and conversion rate
4	Take actions	Action	Conversions and conversion rate

The colors of outdoor advertising design should be relatively vivid and harmonious. The copy should be concise and straight to the center. The font size should be chosen according to the actual size of the outdoor advertisement to avoid reading barriers. Whether it is a static outdoor advertisement or a dynamic outdoor advertisement, its structure should be reasonably laid out according to its function. Refine the structure of outdoor advertising to create a more reasonable outdoor advertising design. In most outdoor advertising designs, patterns dominate the outdoor advertising design.

When conducting outdoor advertising evaluation, we have to analyze specific problems. Different forms of outdoor advertising should be studied using different evaluation methods. We can define some parameters uniformly, but some require specific analysis. When classifying the elements, we can borrow the AHP method. After collecting the index information, we can calculate it by the following equation:

$$P(x) = \sum_{k=1}^{K} \frac{x_k}{K}. \tag{7}$$

where $p(x)$ refers to the average value of primary indicators refers to the sum of primary indicators under the condition that the lower bound is k and the upper bound is K; K refers to the total number of primary indicators; refers to the average value of secondary indicators under the condition that the lower bound is h and the upper bound is H; H refers to the total number of secondary indicators according to the evaluation index system assessment principle, as shown in the following equation:

$$\sum_{k=1}^{K} x_k = \sum_{h=1}^{H} x_h. \tag{8}$$

When making a design, we have some requirements for the proportion of lightness of colors. The ratio of light, dark, and gray tones can directly affect the change of rhythm and expression of emotion in outdoor advertising design. According to research, the proportion of light parts in general outdoor advertising design should not exceed 10% of the total picture. In our creation, high brightness color blocks are generally used to highlight the image, so they should not be excessive. However, for some products, we will use high brightness base colors or even blank base colors to create the design.

Outdoor advertising, whether commercial or public service advertising, is an act of a social nature. Outdoor advertising is public and must maintain the correct direction of its propaganda values while promoting products. When conducting outdoor advertising activities, we must pay attention to the factors of values, moral orientation, aesthetic orientation, and other aspects of spiritual civilization construction in advertisements. The social effectiveness index of outdoor advertising is a comprehensive, long-term, and macroscopic index. The observation of the social effect of outdoor advertising should be continuously observed during the survival of outdoor advertising. Studies with social aspects are complex, so social effect indicators are difficult to be observed from specific values. In determining social effect indicators, we should also consider various complex factors of society.

4. Analysis of Results

4.1. Analysis of Optimization Methods. After calculating the style matching degree using the optimization model, each image has a matching measure value for each of the 15 style terms. The highest scoring font style is used as the corresponding font style category, and the one with the highest match measure value among the 15 style terms is also used as the style feature of the corresponding print advertisement image. Figure 2 shows the statistical distribution of the number of graphic images under each style label. From the results, we can see that the number of images with the "2" style is higher than the number of images with other styles. The number of images with the "12" style is smaller. The uneven distribution of the number of images with different color features is due to the insufficient sample size of the calculation and the difficulty of the labelers to recognize the 15 style terms (Figure 2).

To address the latter, the correlation between the individual style terms was calculated and obtained as shown in Figure 3. In the figure, the red and blue depth of the color indicates the magnitude of the correlation coefficient. The more the color between two style words tends to red, the closer the correlation coefficient is to 1, and the more difficult it is for users to distinguish the difference between them. The more the color between two style words tends to blue, the more the correlation coefficient is close to -1, and the style difference between the two words is obvious. When the correlation coefficient tends to be white, the correlation coefficient is close to 0, indicating that there is no significant correlation between the words. From the results, we can see that the correlation between "luxury" and "glamorous" is 0.712, indicating that the correlation between them is very high, and it is difficult for users to distinguish the degree of difference between these two styles of words. Similarly, the correlation between "chic" and "modern," "classical" and "elegant," "casual" and "elegant," and "casual" and

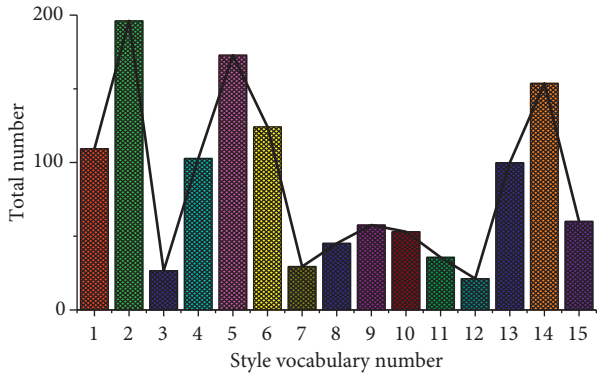


Figure 2: Statistical chart of style characteristics.

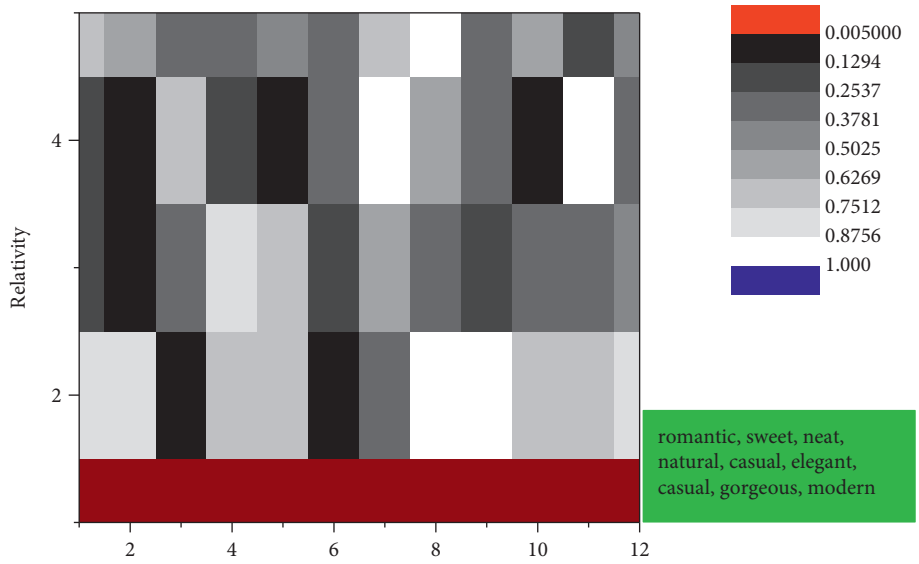


Figure 3: Style keyword relevance settlement results.

"casual" was greater than 0.55. Therefore, the above four pairs of style words were combined, and the style features in print advertising images were finally quantified based on nine style feature labels. These tags are romantic, sweet, neat, natural, casual, elegant, casual, gorgeous, and modern (Figure 3).

4.2. Quantitative Evaluation Analysis. To improve the convergence effect of the distribution, the color characteristics of the background in the text elements can be used as conditions to construct the corresponding conditional probability density functions to estimate the probability distribution of the text color characteristics. Figure 4 shows the distribution of text color brightness under the conditions of different background brightness features. From the results, it can be seen that the text color will be more inclined to high brightness under the background with darker color; under the background with brighter color, the text color brightness is mostly below 1; when the background brightness is in the middle region, not only the text color with high brightness can be used with high probability, but also some low brightness can be used appropriately.

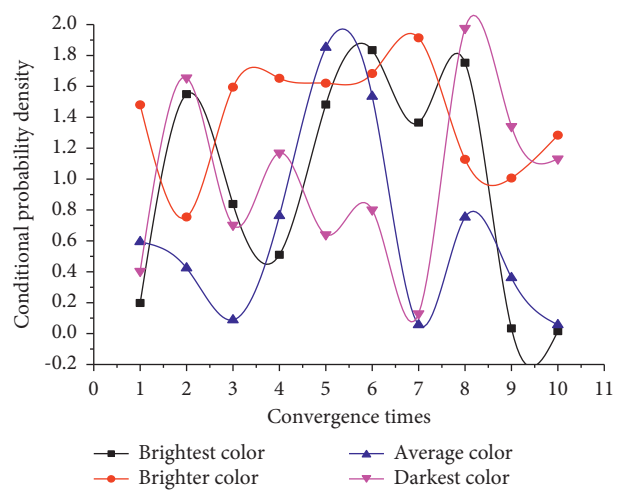


Figure 4: Kernel density estimation and conditional probability density estimation results.

Compared with the kernel density estimation method, some features of graphic design images will have more convergence under conditional probability estimation. The more

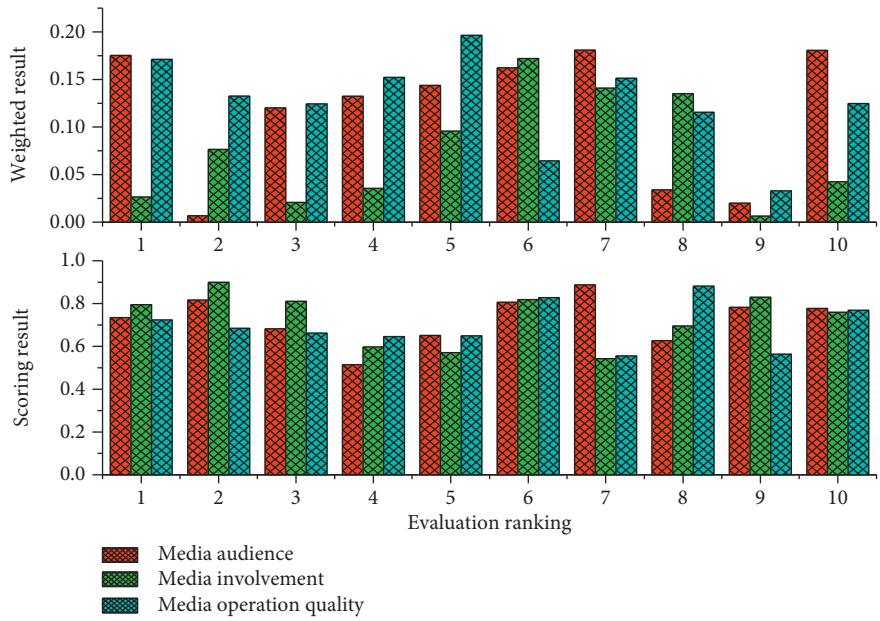


Figure 5: Ranking of indicator classification weights and their scores.

convergent results are also more design-guiding for the computer (Figure 4).

The global weights of all indicators were put together for ranking, and the scores were likewise put together for a comprehensive analysis, as shown in Figure 5. The highest global weight of audience experience is 0.22, and the lowest weight is education level, which accounts for 0.03 of the global weight, which shows that the education level of the media audience does not necessarily have a significant impact on the advertising performance, but the experience of the media audience significantly affects the advertising performance. The global weights of media acceptance, media reach, and property maintenance are 0.21, 0.12, and 0.06, all of which are higher than 0.25 and have a great impact on the performance of community outdoor media advertising. They are indicators that have a great impact on the performance of community outdoor media advertising, and their effective control can improve the operational efficiency of media. Ordinary outdoor billboards: billboards without any lighting equipment belong to the early forms of advertising media. With the continuous development and renewal of outdoor media forms, they will be gradually eliminated (Figure 5).

Through the analysis and evaluation of the index system, we also found that the emotion-driven nature of the images is one of the factors that play a significant role in media involvement. The media has already paid attention to this aspect in its current advertising but still needs to be strengthened in the future in terms of image selection and customer selection. Emotion-driven images are the first step to make consumers resonate with the ads they see and therefore affect several factors such as message interactivity and message relevance, which are factors that must be considered in the process of controlling the effectiveness of outdoor media advertising.

4.3. Optimization Effect Analysis. The decrease in attention to outdoor advertising and the decline in attention to outdoor advertising are considered from both audience and advertiser perspectives. The increase in the number of Internet users and the statistics on the length of time spent online indirectly reflect the change in the focus of the audience, who have shifted more time and energy to the Internet and cell phone media. The characteristics and superiority of the Internet and mobile Internet are reflected in the number of Internet users, related business growth, and profound changes in people's living conditions. In 2020, the number of Internet users was 800 million, with a penetration rate of 51.37%, an increase of 1.23% over the previous year, and the number of cell phone users was 656 million, and the total number of Internet users and cell phone users in China showed a rising trend. In the first half of the year, the weekly Internet access hours per capita was 27.3 hours, an increase of 0.4 hours compared with 2019 (see Figure 6).

With the expansion of Internet users and the increase of Internet users' time on the Internet, advertisers naturally also see this trend, and they will consider the value of media and pay more attention to the marketing effect of advertising and maximizing the value of information dissemination while reducing advertising expenditures, and advertisers begin to put their eyes on new media platforms to obtain effective dissemination of information and timely feedback. According to the core data of online advertising in 2016, the market size of online advertising reached 3984.3 billion yuan, with a year-on-year growth of 36.87%. The market size of mobile advertising reached 2145.1 billion yuan in 2016, with a year-on-year growth rate of 157.3%, and the development momentum was very strong (see Figure 7).

The experiment recruited 100 people to participate in a subjective questionnaire. The questionnaire presented

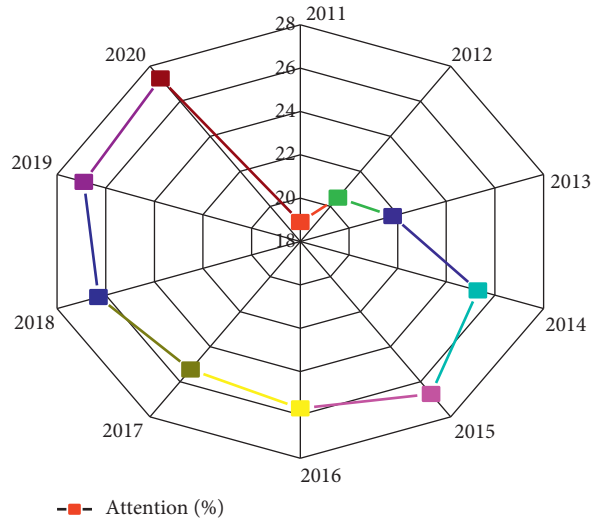


Figure 6: The trend of average Internet users' time spent online.

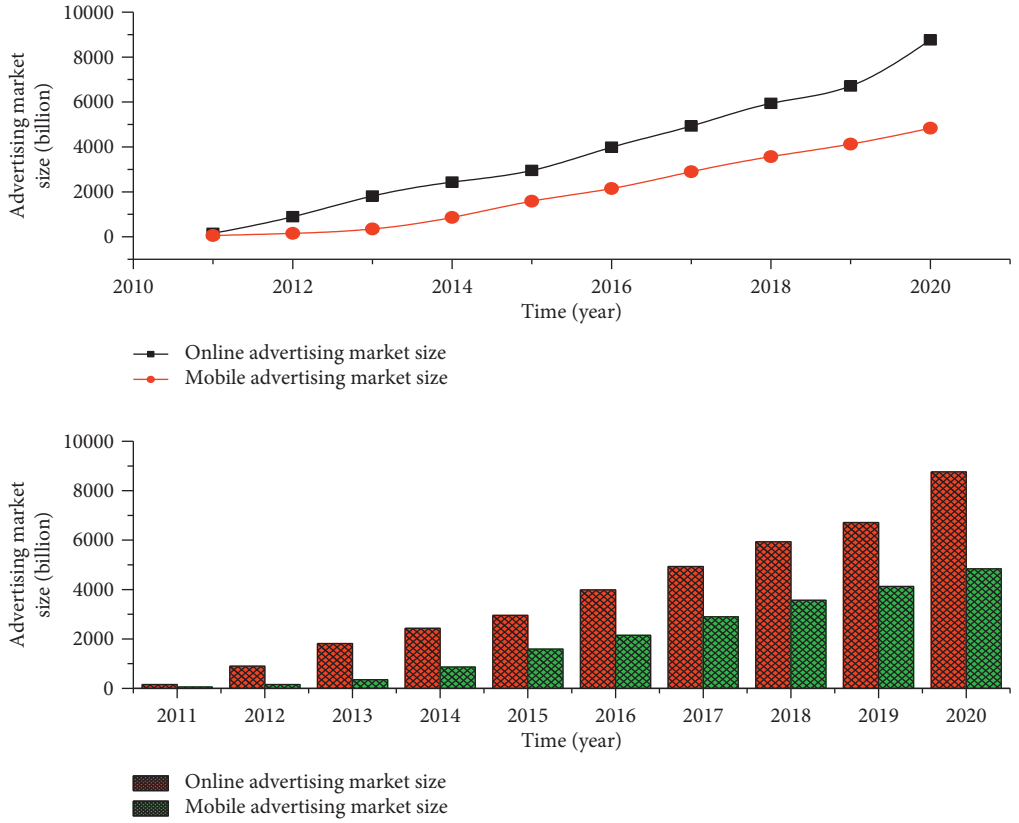


Figure 7: Outdoor advertising market size and forecast.

participants with 10 color outdoor advertisements of the same layout template each time and asked participants to select 5 outdoor advertisements that they thought had the best color design and 5 outdoor advertisements that had the worst color design. The design results of one of the layout templates were presented twice during the questionnaire to check the consistency of participants' answers. Figure 8 shows the results of the user selection for different color feature models. In this case, red indicates the number of designed outdoor advertisements that users consider to be the best, blue indicates the number of designed outdoor advertisements that they consider to be the worst, and the number of other outdoor advertisements is indicated in yellow (Figure 8).

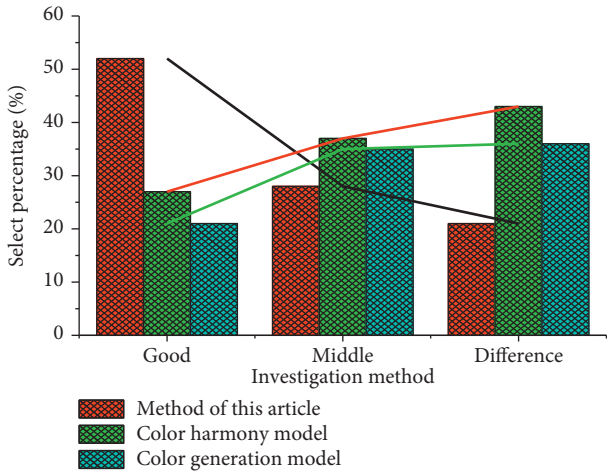


Figure 8: Outdoor advertising optimization results.

The social effect produced by new media interactive advertising is mainly the influence of advertising activities on social morality and cultural education. It is difficult to quantify the social effect of new media interactive advertising with a few specific indicators like the evaluation of communication effect and economic effect, because the evaluation of social effect involves all aspects of the whole social environment such as political philosophy, economic form, legal norms, ethics, and morality, which cannot be expressed in specific indicators and values and can only be macroscopically grasped through several aspects such as social influence, buzz, and reputation. The only way to grasp them is through social influence, popularity, and reputation. Similarly, when assessing the social effect of new media interactive advertising, it is also subject to the influence and constraints of social culture in the corresponding social form. Different ideologies and social environments will lead to different results of social evaluation.

5. Conclusion

With the development of science and technology, the form of advertising has also ushered in diversified changes. In terms of advertising effects, big data interactive advertising under the sieve of mature data collection, processing, and analysis system avoids the inefficient propaganda mode of traditional advertising indiscriminate casting and starts from the needs of the audience, putting the practicality and entertainment of advertising in the first place. Based on the exploration and research on spiritual civilization, this paper puts the main research direction on the exploration of subjective elements. Starting from the designer's perspective and observing from the audience's perspective, we get some elements that are more in-depth than the original research, making the whole system more fleshed out and the research more comprehensive. Unlike the passive way of audience reception in traditional advertising, new media interactive advertising establishes a more direct way of communication between consumers and products, creates different five senses stimulation and emotional experience environment according to the characteristics of products, and carries out a series of interactions with consumers' sensory systems such as vision, hearing, smell, taste, and touch, which touches consumers' emotional factors, and the way of advertising communication has changed from passive reception to active participation. From the perspective of advertising, the change of advertising communication is interpreting the arrival of a new communication concept, the Internet-centered, and technology-centered data integration era, and we are gradually seeing the huge changes brought by technology and human-centered design to human society and life. Outdoor advertising is facing the impact of new media, on the one hand, to maintain and play its media value and, on the other hand, to actively respond from media competition to media integration. The unique value of outdoor advertising is the intrinsic reason to support its development, and the integration and complementarity between media provide external support for outdoor advertising. I believe that, after the development of the times and the improvement of technology, the outdoor media market will have a better development space in the future.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

No conflicts of interest exist concerning this study.

References

[1] I. Mademlis, V. Mygdalis, N. Nikolaidis et al., "High-level multiple-UAV cinematography tools for covering outdoor events," *IEEE Transactions on Broadcasting*, vol. 65, no. 3, pp. 627–635, 2019.

[2] C. Campbell, S. Sands, C. Ferraro, H.-Y. Tsao, and A. Mavrommatis, "From data to action: how marketers can leverage AI," *Business Horizons*, vol. 63, no. 2, pp. 227–243, 2020.

[3] B. Boştină-Bratu, A. Negoescu, and L. Palea, "Consumer acceptance of outdoor advertising: a study of three cities," *Land Forces Academy Review*, vol. 23, no. 1, pp. 65–74, 2018.

[4] S. Chua and A. Duffy, "Friend, foe or frenemy? Traditional journalism actors' changing attitudes towards peripheral players and their innovations," *Media and Communication*, vol. 7, no. 4, pp. 112–122, 2019.

[5] U. Kose and P. Vasant, "Better campus life for visually impaired University students: intelligent social walking system with beacon and assistive technologies," *Wireless Networks*, vol. 26, no. 7, pp. 4789–4803, 2020.

[6] Y. Choi, B. Hickerson, and J. Lee, "Investigation of the technology effects of online travel media on virtual travel experience and behavioral intention," *Journal of Travel & Tourism Marketing*, vol. 35, no. 3, pp. 320–335, 2018.

[7] J. Sproul, S. Ledger, and J. MacCallum, "A review of digital media guidelines for students with visual light sensitivity," *International Journal of Disability, Development and Education*, vol. 68, no. 2, pp. 222–239, 2021.

[8] P. Ozcan and D. Hannah, "Forced ecosystems and digital stepchildren: reconfiguring advertising suppliers to realize

disruptive social media technology," *Strategy Science*, vol. 5, no. 3, pp. 193–217, 2020.

[9] U. Stankov, J. Kennell, A. M. Morrison, and M. D. Vujičić, "The view from above: the relevance of shared aerial drone videos for destination marketing," *Journal of Travel & Tourism Marketing*, vol. 36, no. 7, pp. 808–822, 2019.

[10] E. Karaaslan, U. Bagci, and F. N. Catbas, "Artificial intelligence assisted infrastructure assessment using mixed reality systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 12, pp. 413–424, 2019.

[11] X. Chen, J. Chen, G. Cheng, and T. Gong, "Topics and trends in artificial intelligence assisted human brain research," *PLoS One*, vol. 15, no. 4, p. e0231192, 2020.

[12] Y. Feng, Y. Zhao, and H. Zheng, "Data-driven product design toward intelligent manufacturing: a review," *International Journal of Advanced Robotic Systems*, vol. 17, no. 2, p. 1729881420911257, 2020.

[13] Y. Liu, "Evaluating visitor experience of digital interpretation and presentation technologies at cultural heritage sites: a case study of the old town," *Built Heritage*, vol. 4, no. 1, pp. 10–15, 2020.

[14] Q. Xie, E. Schauster, and M. S. Neill, "Expectations for advertising and public relations education from agency executives: a comparative study between China and the United States," *Journal of Current Issues and Research in Advertising*, vol. 39, no. 3, pp. 289–307, 2018.

[15] H. Lee and C.-H. Cho, "Digital advertising: present and future prospects," *International Journal of Advertising*, vol. 39, no. 3, pp. 332–341, 2020.

[16] G. Monardo, C. Pavese, I. Giorgi, M. Godi, and R. Colombo, "Evaluation of patient motivation and satisfaction during technology-assisted rehabilitation: an experiential review," *Games for Health Journal*, vol. 10, no. 1, pp. 13–27, 2021.

[17] G. Rosenkrans and K. Myers, "Optimizing location-based mobile advertising using predictive analytics," *Journal of Interactive Advertising*, vol. 18, no. 1, pp. 43–54, 2018.

[18] W. Nawaz, K. U. Khan, and K. Bashir, "A review on path selection and navigation approaches towards an assisted mobility of visually impaired people," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 14, no. 8, pp. 3270–3294, 2020.

[19] N. A. F. M. Kamarolzaman, C. F. Yeong, and E. L. M. Su, "Smart advertising robot with image recognition for data analytics," *ELEKTRIKA-Journal of Electrical Engineering*, vol. 16, no. 3, pp. 11–16, 2017.

[20] C.-C. Huang, "User's segmentation on continued knowledge management system use in the public sector," *Journal of Organizational and End User Computing*, vol. 32, no. 1, pp. 19–40, 2020.

[21] N. Ramu, V. Pandi, J. D. Lazarus, and S. Radhakrishnan, "A novel trust model for secure group communication in distributed computing," *Journal of Organizational and End User Computing*, vol. 32, no. 3, pp. 1–14, 2020.

[22] A. Aderonke, "Oni, ugbedeojo musa, samuel oni. E-Revenue adoption in state internal revenue service: interrogating the institutional factors," *Journal of Organizational and End User Computing*, vol. 32, no. 1, pp. 41–61, 2020.

[23] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 577–590, 2018.

[24] N. Alvertos, D. Brzakovic, and R. C. Gonzalez, "Camera geometries for image matching in 3-D machine vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 09, pp. 897–915, 1989.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 1336900, 11 pages
https://doi.org/10.1155/2021/1336900

Research Article

Missing Data Reconstruction Based on Spectral *k*-Support Norm Minimization for NB-IoT Data

Luo Xuegang,[1] **Lv Junrui**,[1] **and Wang Juan**[2]

[1]*School of Mathematics and Computer Science, Panzhihua University, Panzhihua 617000, China*
[2]*College of Computer Science, China West Normal University, Nanchong 637000, China*

Correspondence should be addressed to Wang Juan; wjuan0712@126.com

Received 23 September 2021; Accepted 16 October 2021; Published 10 November 2021

Academic Editor: Xianyong Li

An effective fraction of data with missing values from various physiochemical sensors in the Internet of Things is still emerging owing to unreliable links and accidental damage. This phenomenon will limit the predicative ability and performance for supporting data analyses by IoT-based platforms. Therefore, it is necessary to exploit a way to reconstruct these lost data with high accuracy. A new data reconstruction method based on spectral *k*-support norm minimization (DR-SKSNM) is proposed for NB-IoT data, and a relative density-based clustering algorithm is embedded into model processing for improving the accuracy of reconstruction. First, sensors are grouped by similar patterns of measurement. A relative density-based clustering, which can effectively identify clusters in data sets with different densities, is applied to separate sensors into different groups. Second, based on the correlations of sensor data and its joint low rank, an algorithm based on the matrix spectral *k*-support norm minimization with automatic weight is developed. Moreover, the alternating direction method of multipliers (ADMM) is used to obtain its optimal solution. Finally, the proposed method is evaluated by using two simulated and real sensor data sources from Panzhihua environmental monitoring station with random missing patterns and consecutive missing patterns. From the simulation results, it is proved that our algorithm performs well, and it can propagate through low-rank characteristics to estimate a large missing region's value.

1. Introduction

An Internet of Things (IoT) platform including sensors, wireless communication, and data processors has been deployed to support remote monitoring and intelligent analysis. At present, a large number of smart industries have emerged, such as smart health, smart campus, smart finance, smart retail, and smart agriculture, which benefit from the rise of the Internet of Things technology. Among them, narrowband IoT (NB-IoT) technology is developing at a rapid speed. Compared with the 3G network, 4G network, wireless sensor network (WSN), Wi-Fi, low-power wide-area network from the LoRa Alliance (LoRaWAN), Zigbee, and so on, it has the characteristics of wide coverage, multiple connections, low rate, low cost, low-power consumption, and excellent architecture [1]. Therefore, NB-IoT is rapidly transforming into a highly heterogeneous ecosystem that provides information exchange among different types of information sensor equipment and communications technologies.

Environmental monitoring platform based on NB-IoT, which provides monitoring, analysis, governance, and protection of environmental quality for relevant government departments, solves the shortcomings of single environmental monitoring, complex wiring, poor transmission capacity, time-consuming, and laborious [2, 3]. IoT has played an important role to strengthen national competitiveness and attracted the attention of industry, academia, government, and regulators worldwide [4]. A large number of IoT platforms have been developed for monitoring, supervision, and management. In addition to collecting data by sensors, IoT platforms also need to have the function of intelligent analysis by big data and artificial intelligence technology. Because these

data are large-scale, high-dimension, and complex structures, there are strict requirements in terms of data integrity, correctness, and on-time delivery. However, in many IoT applications, especially the environmental monitoring of the field environment, it is unavoidable that massive data are lost. On the one hand, owing to unreliable links and accidental damage, data loss or damage may occur during acquisition, transmission, and storage. On the other hand, the limited capability of sensor nodes impacts data collection in terms of energy, storage, and communication. Some issues with battery depletion and device failure result in imperfect data gathering.

Incomplete data collected by sensors make it more difficult to process sensor data, hindering various high-level applications of the Internet of Things, such as intelligent analysis and prediction. If missing data values cannot be accurately recovered, existing intelligent analysis tools cannot be applied. If lost data are deleted directly, a large amount of original data becomes unavailable, which reduces the accuracy and reliability of analysis results and wastes a large amount of data resources. The amount of data collected by the Internet of Things every day around the world is enormous. Currently, that is about 26 billion bytes per day, and the number will increase dramatically. Recovering lost sensor data effectively to analyze IoT applications accurately is a major challenge. Designing effective methods to reconstruct loss sensor data has a wide range of applications. Therefore, it is very urgent and important to design an effective method to reconstruct the missing values in sensor data.

Many methods of populating missing sensor data have been developed by researchers. Methods based on time, space, or a combination of space and time are predominant. However, in some special cases, high-frequency data loss may obscure the temporal and spatial correlation between data. Therefore, it is necessary to study new methods to solve high-frequency data loss. As is known to all, a sensor node usually has multiple sensors integrated. These nodes usually collect multiple monitoring data at the same time, and the data collected by these nodes have a certain correlation. Therefore, it is beneficial to improve the estimation accuracy in predicting sensor data, which can use the correlation between different types of data as a supplement to the internal correlation.

In this paper, a new data reconstruction method based on spectral k-support norm minimization [5] is proposed for NB-IoT data. This method implements the relative density-based clustering algorithm [6] to separate sensors into different groups. The main objective of clustering is to increase the similarity within the same group in terms of the spatial relationship of sensor nodes. Thereafter, we use the correlations of sensor data and its joint low rank; an algorithm based on the matrix weighted Schatten-p norm minimization with automatic weight for filling the multiattribute sensor data within each cluster is developed. Moreover, the alternating direction method of multipliers (ADMM) is used to obtain its optimal solution. Our contributions are summarized as follows:

(1) Considering the uneven distribution of sensor nodes, the relative density-based clustering algorithm is applied to divide the sensor nodes into different clusters according to the distribution around neighbor points of the data points, to ensure that similar patterns of measurement of sensors are within one group

(2) The data reconstruction method based on spectral k-support norm minimization is applied for reconstruction of missing sensor data, taking advantage of the low-rank feature between different attributes of sensor data

The remainder of this paper is organized as follows. In Section 2, background and related work are presented. Section 3 describes our proposed method. The performance of the proposed method is evaluated in Section 4. Section 5 provides our concluding remarks and a simple elaboration for future work.

2. Background and Related Work

NB-IoT, which is a wireless telecommunications technology standard, is developed to connect multiple Internet of Things devices and empower IoT architecture using existing mobile networks. The NB-IoT network has been widely employed in industry, agriculture, healthcare, and logistics, and, quite obviously, in smart cities and buildings. Its purpose is to facilitate the work of technicians, retailers, and operators while handling machines by providing real-time technological data and supporting remote monitoring systems. According to recent studies, by 2026, cellular LPWAN solutions (that is, NB-IoT and LTE-M combined) will be responsible for over 60% of the estimated 3.6 billion low-power wide-area network connections, with the remaining 40% covered by noncellular, of which LoRa and Sigfox will account for the majority. Figure 1 shows NB-IoT end-to-end system architecture including sensor nodes, NB-IoT wireless networks, application server platform, and user terminal equipment. Many techniques in the literature based on temporal methods, spatial methods, and machine learning-based methods have focused on loss data prediction in NB-IoT.

Temporal methods, namely, the mean of observation data [7], the last observation and linear interpolation, and so on, exploit the time correlations between the readings recorded by the same sensor node. Because the data obtained by the sensor node near the monitoring range often varies slowly over a short time period. However, when the observation interval of a given sensor is long or the sensor network changes drastically and irregularly, the temporal methods do not work well, and as the number of consecutive missing readings increases, its effectiveness decreases rapidly.

Spatial methods utilize the spatial correlation between the sensor data of spatially similar sensor nodes at the same monitoring time; generally, the closer the sensor nodes are in space, the greater the correlation of the data is. The missing

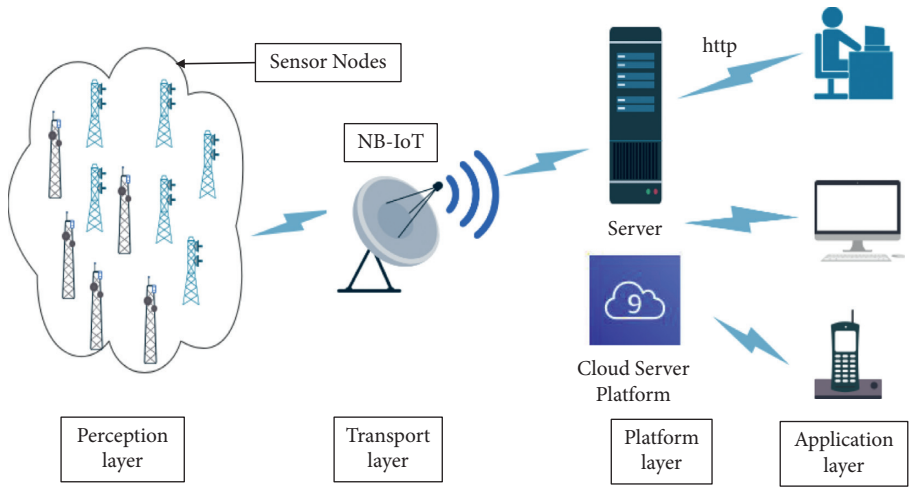


Figure 1: NB-IoT end-to-end system architecture.

data estimation algorithm based on *K*-nearest neighbor [8] takes advantage of the values of the nearest K neighbors to estimate the missing one. Rahman et al. [9] have described a missing data imputation approach that was combining Fourier and lagged k-nearest neighbor for biomedical time series missing data imputation. The nearest neighbor (NN) imputation method that estimates missing data in WSNs by learning spatial and temporal correlations between sensor nodes was proposed by Li and Parker [10]. An MP-BMDI algorithm for univariate time series with large missing gap was proposed by Lee [11]. However, such models, which typically require precise NN distances between sensors, are not suitable only for large missing gap of signals and environments within univariate time series.

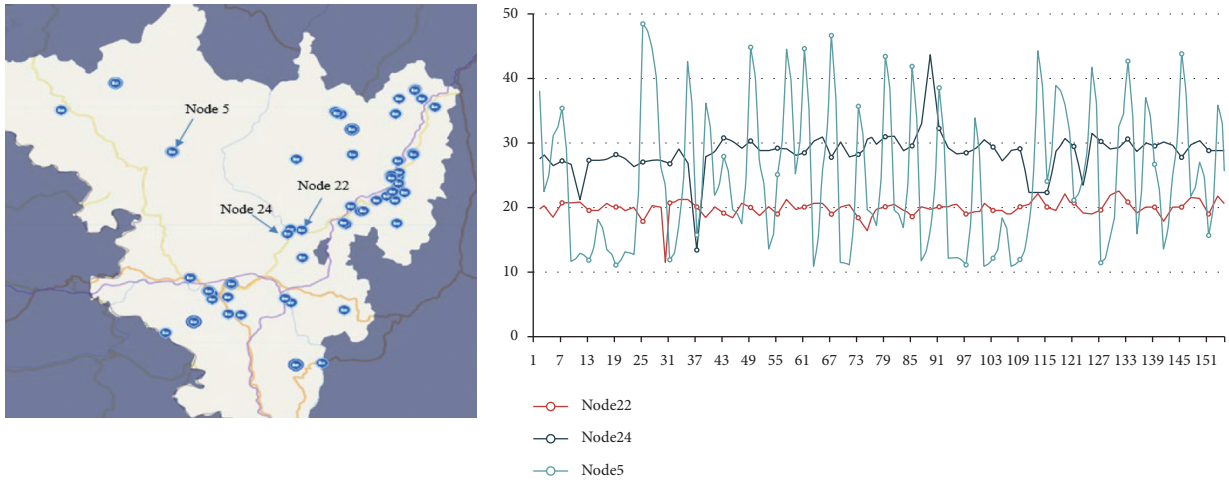
Machine learning-based methods utilizing the machine learning techniques mainly focus on the data correlation structure of the whole dataset to estimate the missing values [12–15]. Here, the multiple linear regression model (MLR) [12] using spatio-temporal correlation was proposed for missing data reconstruction of WSN Data. Zhang and Yin [13] investigated a multivariate time series missing data imputation by recurrent denoising autoencoder and showed legitimate reconstruction performance in a realistically high-dimensional dataset. In [14], the authors presented a novel estimation method of multivariate time series missing data by adapting the bidirectional long short-term memory (LSTM). Pattern sequence forecasting was applied to reconstruct loss data in [15]. The performance of machine learning-based methods depended on accurate historical time series data, but those data are often incomplete and missing.

All the above methods aiming to estimate missing values suffered from over-relying on assumptions about data. For example, when the sampling frequency of the sensor decreases, the usefulness of temporal methods may drop rapidly. Besides, estimation results may be worse as nonexistent spatial correlations are imposed between nearby sensors. In the same way, a priori knowledge that nonexistent temporal correlation was favored for missing data imputation will lead to performance degradation. Therefore, it is necessary to directly exploit sensor latent structures from data without heavily relying on assumptions. Recently, the intrinsic low-rank property of high-dimensional data has been applied to exploit latent structures of sensor data. Weighted nuclear norm minimization [16] was exploited to multiattribute missing data reconstruction for improving IoT data reliability. Despite numerous methods have been developed for imputing missing values, imputation precision and computational complexity are still the major issues for large missing subsequences. To solve this problem, a newly developed algorithm utilized weighted spectral k-support norm minimization that ensures high reconstruction performance for supporting IoT-based applications and data analysis in this paper when large missing subsequences are absent.

3. Our Proposed Method

In this section, the relative density-based clustering applied to separate sensors into different groups is presented; then, a detailed aspect of our proposed reconstruction procedure that tensor composed of sensor multiattribute data is recovered by weighted spectral *k*-support norm minimization is described.

3.1. Sensor Nodes Grouping with the Relative Density-Based Clustering Algorithm. Generally, many sensor nodes are deployed in a monitored region and have spatial correlations from [8]. For example, Figure 2(a) shows the sensor nodes locations from Panzhihua city in China and marked three nodes, among which sensor node 22 and sensor node 24 are close to each other, while sensor node 5 is far away. The deployment locations of sensor nodes are unknown for monitoring purposes. In Figure 2(b), chemical oxygen demand (COD) (unit: mg/L) values of sensor node 5, node 22, and node 24 are curve plotted from the urban environmental water quality monitoring dataset in Panzhihua. Figure 2(b) demonstrates that the trend of monitoring index curve remained consistent of sensor node 5 and sensor node 22.



(a) (b)

FIGURE 2: Sensor nodes with monitoring index location. (a) Sensor nodes locations in a region. (b) COD (unit: mg/L) collected by three sensor nodes.

However, the index curve of sensor node 5 is far from nodes 22 and 24 as monitoring is characterized by concentration and aggregation. Neighboring nodes have similar attributes and similar data changes. Thus, when some of the sensed data of a sensor node are missing, we can estimate them by using the data of the neighboring nodes. This example indicated that not every node in the monitoring region is useful for recovering the missing values. Thus, it is necessary to first divide the sensors into different groups to minimize measurement changes within each group for utilizing similarity well among measured values of neighboring sensors.

The K means algorithm, which is the most typical clustering method [17], can effectively recognize convex clusters. However, K means cannot precisely identify the true clusters for nonconvex clusters. The density-based clustering method can find clusters of various shapes and sizes in data with noisy [6]. The relative density-based clustering algorithm can effectively find clusters in data sets with nonuniform density.

In this work, we apply the relative density-based clustering algorithm to group sensors. This algorithm divides a set of points into many clusters in terms of the relative density of each point and distance so that points in each cluster tend to be close to each other.

Given the data set P and the k-distance of point d, the k -distance neighbors of point d are points whose distance from d is not greater than the k -distance, defined by

$$D_{p\_dist(d)}(d) = \{d \in P \mid dis(b,d) \le k_{\text{distance}}(d)\}. \quad (1)$$

Reachability distance of point d with respect to point s is defined as

$$\text{reach}_{\text{dist}(d,s)} = \max\{k_{\text{distance}}(s) dis(d,s)\}. \quad (2)$$

where $dis(d,s)$ denotes the distance of point d and point s and $k_{\text{distance}}(s)$ is the k-nearest neighbors.

The local reachability density of d is the inverse of the average based on the k-nearest neighborhood of a point d, defined as

$$\text{reach\_dsity}_k(d) = \frac{1}{\left(\sum_{s \in D_k(d)} \text{reach}_{\text{dist}(d,s)}\right) / \left|D_k(d)\right|}. \quad (3)$$

We simplify the notation and use $D_k(d)$ as a shorthand for $D_{p\_dist(d)}(d)$.

Then, the relative density of d with respect to its $D_{p\_dist(d)}(d)$ neighbors denoted as $\text{rel\_dsity}_k(d)$ is defined as

$$\text{rel\_dsity}_k(d) = \frac{\sum_{s \in D_k(d)} \text{reach\_dsity}_{k(s)} / \text{reach\_dsity}_{k(d)}}{\left|D_k(d)\right|}. \quad (4)$$

The main steps of the relative density-based clustering algorithm are as follows:

Step 1: set data set of points P, and initialize the clustering center node-set $C = \{0\}$

Step 2: calculate the distance between point j and point r, and then obtain dist[j, r]

Step 3: for each point, calculate the k-distance neighbors and calculate the k-distance neighbors, reachability distance, local reachability density, and relative density according to (1)–(4)

Step 4: select points whose relative density is less than 1 and store them in set IC, and C is filled with assignment method (IC, dist[j, r]) (see [6])

Step 5: for each point d in data set P, group them to the nearest cluster C_i

Step 6: iterate steps 2, 3, 4, and 5 until set C no longer changes or the difference between adjacent iterations is smaller than the given threshold ε

After the above algorithm, sensors of the urban environmental water quality monitoring dataset in Panzhihua are divided into five groups (see Figure 3).

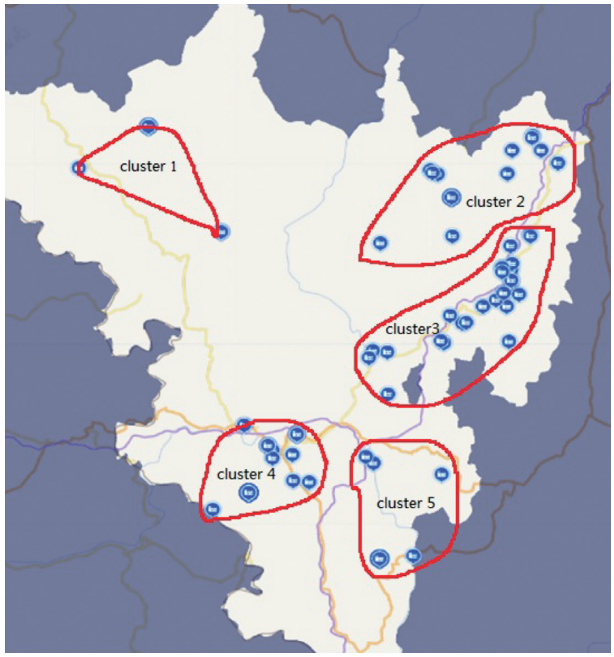


Figure 3: Clustering result of sensor location in the Panzhihua region.

3.2. Data Reconstruction with Spectral k-Support Norm Minimization

3.2.1. Spectral k-Support Norm. As sensor data have a low-rank characteristic, the problem of missing data reconstruction can be regarded as the rank minimization problem. The goal is to recover missing values of a matrix $M \in \mathbb{R}_r^{m\times n}$ from partial observations, supposing that its rank $r \ll \min\{m,n\}$ among all matrices with the same observations, and m and n are the number of rows and columns of the matrix. The following model for the solution of the rank minimization problem can estimate the unknown M:

$$\min_{\mathbf{X}} \operatorname{rank}(X), s.t. \|X - M\|_{\mathrm{F}}^2 < \delta, X_{ij} = M_{ij}, \quad \forall (i,j) \in \Omega, \tag{5}$$

where Ω is the set of the observed indices; $\delta > 0$ is a small constant; and $\|.\|^{\mathrm{F}}$ represents the norm. It is difficult to solve problem (5), that is, NP-hard [18]. To obtain a solution in polynomial time, the nuclear norm is proposed as a substitute. The nuclear norm-based matrix completion model is formulated as follows:

$$\min_{\mathbf{X}} \lambda \|\mathbf{X}\|_*, s.t. \|X - M\|_{\mathrm{F}}^2 < \delta, X_{ij} = M_{ij}, \quad \forall (i,j) \in \Omega, \tag{6}$$

where $\|X\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(X)$ is the kernel norm, λ is a constant and greater than zero, and $\sigma_i$is the ith singular value. Problem (6) adopts singular value threshold contraction of the observation matrix, namely,

$$\begin{cases} M = U\Sigma V^T, \\ X = U\zeta_\lambda(\Sigma)V^T, \end{cases} \tag{7}$$

where $U\Sigma V^T$ is SVD decomposition, $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{r\times n}$ are the decomposed orthogonal matrix, $\zeta_\lambda(\Sigma)_{ii} = \max(\Sigma_{ii} - \lambda, 0)$ is the soft threshold function of a diagonal matrix Σ, and Σ_{ii} is the iterative solution of diagonal matrix elements in (6) until convergence.

Recently, the spectral k-support norm [19] has been shown to have good estimation properties in low-rank matrix learning problems and better captures the spectral decay of the underlying model, and the spectral k-support norm-based matrix completion model is as follows:

$$\min_{\mathbf{X}} \|X - M\|_{\mathrm{F}}^2 + \|\mathbf{X}\|_{(k)}^{msp} \text{ s.t. } X_{ij} = M_{ij}, \quad \forall (i,j) \in \Omega. \tag{8}$$

where $\|X\|_{(k)}^{msp}$denotes spectral k-support norm. The spectral k-support norm is the gauge function whose unit ball is the convex set:

$$\operatorname{conv}\{\sigma | \|\sigma\|_0 \le k, \|\sigma\|_F \le 1\}, k \in \mathbb{N}_+. \tag{9}$$

Thus, the unit ball of the spectral k-support norm is the convex hull of matrices, and its rank is no larger than k and Schatten-p norm no larger than 1. The spectral k-support norm-based model has shown superior performance over the nuclear norm-based models.

It can be the following explicit computation:

$$\|X\|_{(k)}^{msp} := \|\sigma\|_{(k)}^{msp} = \left[\sum_{i=1}^{k-l-1} (\sigma_i)^2 + \frac{1}{l+1}\left(\sum_{j=k-l}^{D} \sigma_j\right)^2\right]^{1/2}, \tag{10}$$

where $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_{\min(m,n)}$ are the singular values of Min nonincreasing order and l is the unique integer in $[0, k-1]$ satisfying

$$\sigma_{k-l-1} \ge \frac{1}{l+1} \sum_{j=k-l}^{\min(m,n)} \sigma_j \ge \sigma_{k-l}. \tag{11}$$

3.2.2. Data Reconstruction Algorithm and ADMM-Based Solution. The sensor data gathered in one node can be organized by the following format of sensor ID standing for sensor identity number, timestamp representing sampling time, and the monitoring index parameters including chemical oxygen demand, ammonia nitrogen, residual chlorine, and humidity.

Let $\mathcal{T} \in \mathbb{R}^{Q\times P\times C}$ be a tensor, whose data are composed of Q monitoring index parameters collected by P nodes within C time slots. The entry $\mathcal{T}$ is represented by $t_{q,p,c}$. Because of data loss in NB-IoT, $\mathcal{T}$ is usually an incomplete tensor. A set of entries of $\mathcal{T}$ were intact information gathered by sensors. We use $[\mathcal{G}_\Omega(\cdot)]$ to indicate the sampling operator, which is defined as follows:

$$[\mathcal{G}_\Omega(\mathcal{T})]_{q,p,c} = \begin{cases} t_{q,p,c}, \text{if } (q,p,c) \in \Omega, \\ 0, \quad \text{otherwise.} \end{cases} \tag{12}$$

The sensor missing data reconstruction problem with spectral k-support norm minimization is defined as follows:

$$\begin{aligned} &\min \|\mathscr{L}\|_{(k)}^{tsp} + \lambda \|\mathscr{R}\|_1 \\ &s.t. \quad \mathscr{G}_\Omega(\mathscr{T}) = \mathscr{G}_\Omega(\mathscr{L}) \text{ and } \mathscr{H} * (\mathscr{T} - \mathscr{L}) = \mathscr{R}, \end{aligned} \tag{13}$$

where $\|\bullet\|_1$ is the Frobenius norm, $\|\cdot\|_{(k)}^{tsp}$ denotes tensor spectral k-support norm, that is, an extension of matrix spectral k-Support norm $\|\cdot\|_{(k)}^{msp}$, and $\mathscr{H}$ is an identity mapping and the element-wise projection, correspondingly. As the sensor data tensor is low rank, the recovery tensor $\mathscr{L}$ is restricted by tensor spectral k-support norm minimization in (13).

The alternate direction multiplier method (ADMM) algorithm, a convex optimization algorithm, is a good candidate for the low-rank model problem when compared with the other state-of-the-art splitting proximal algorithms. Therefore, in the present study, we attempted to use the ADMM algorithm to solve the formulated problem. To apply the ADMM method to solve equation (13) in the paper via ADMM-type algorithm, adding a new tensor-valued auxiliary variable $\mathfrak{J}$, the augmented Lagrangian is given by

$$\begin{aligned} \mathscr{D}_\beta(\mathscr{L}, \mathscr{R}, \mathfrak{J}) = &\frac{1}{2}\|\mathscr{L}\|_{(k)}^{tsp} + \lambda \|\mathscr{R}\|_1 \\ &+ \langle \mathfrak{J}, \mathscr{H} * (\mathscr{T} - \mathscr{L}) - \mathscr{R} \rangle \\ &+ \frac{\beta}{2}\|\mathscr{H} * (\mathscr{T} - \mathscr{L}) - \mathscr{R}\|_F^2, \end{aligned} \tag{14}$$

where β is a penalty parameter from the augmented Lagrangian formulation and λ is a constant from ℓ_1 norm. In the following, we carry out the alternative update to the variables $\mathscr{L}_t$ and $\mathscr{R}_t$ at iteration t in detail.

(1) Update $\mathscr{L}_t$ as follows:

$$\begin{aligned} \mathscr{L}_t = \arg\min_{\mathscr{L}} &\frac{1}{2}\|\mathscr{L}\|_{(k)}^{tsp} + \langle \mathfrak{J}_{t-1}, \mathscr{H} * (\mathscr{T} - \mathscr{L}) - \mathscr{R}_{t-1} \rangle \\ &+ \frac{\beta}{2}\|\mathscr{H} * (\mathscr{T} - \mathscr{L}) - \mathscr{R}_{t-1}\|_F^2. \end{aligned} \tag{15}$$

To separate $\mathscr{H}$ apart from $\mathscr{L}$, the preconditioned ADMM approximates $1/2\|\mathscr{H} * (\mathscr{T} - \mathscr{L}) - \mathscr{R}_{t-1}\|_F^2$ with second-order Taylor expansion around $\mathscr{L}_{t-1}$ as

$$\frac{1}{2}\|\mathscr{H} * (\mathscr{T} - \mathscr{L}_{t-1}) - \mathscr{R}_{t-1}\|_F^2 - \langle \mathscr{H}^{\mathrm{T}} * (\mathscr{H} * (\mathscr{T} - \mathscr{L}_{t-1}) - \mathscr{R}_{t-1})\mathscr{L} - \mathscr{L}_{t-1} \rangle + \frac{\delta}{2}\|\mathscr{L} - \mathscr{L}_{t-1}\|_F^2. \tag{16}$$

Incorporating equation (16) into equation (15) gives

$$\begin{aligned} \mathscr{L}_t = \arg\min_{\mathscr{L}} &\frac{1}{2\beta\delta}\|\mathscr{L}\|_{(k)}^{tsp} + \langle \mathfrak{J}_{t-1}, \mathscr{H} * (\mathscr{T} - \mathscr{L}) - \mathscr{R}_{t-1} \rangle \\ &+ \frac{1}{2}\left\| \mathscr{L} - \left(\mathscr{L}_{t-1} + \mathscr{H}^T * (\mathscr{H} * (\mathscr{T} - \mathscr{L}_{t-1}) - \mathscr{R}_{t-1}) + \frac{\mathfrak{J}_{t-1}}{\beta\delta} \right) \right\|_F^2. \end{aligned} \tag{17}$$

Thus, according to the proximal operator for TSP-k norm in [5], problem (17) has a closed-form solution introduced to solve this problem. The computation of the proximal map in equation (17) has been given by

$$\mathscr{L}_t = \mathrm{Prox}_{\frac{1}{2\beta\delta}\|\cdot\|_{tsp,k}^2}\left(\mathscr{L}_{t-1} + \mathscr{H}^T * (\mathscr{H} * (\mathscr{T} - \mathscr{L}_{t-1}) - \mathscr{R}_{t-1}) + \frac{\mathfrak{J}_{t-1}}{\beta\delta} \right). \tag{18}$$

(2) Update $\mathscr{R}_t$ as follows:

$$\begin{aligned}\mathscr{R}_t &= \arg\min_{\mathscr{R}} \lambda\|\mathscr{R}\|_1 + \langle \mathfrak{Y}_{t-1}, \mathscr{H} * (\mathscr{T} - \mathscr{L}_t) - \mathscr{R}\rangle + \frac{\beta}{2}\|\mathscr{H} * (\mathscr{T} - \mathscr{L}_t) - \mathscr{R}\|_F^2 \\ &= \arg\min_{\mathscr{R}} \frac{\lambda}{\beta}\|\mathscr{R}\|_1 + \frac{1}{2}\left\|\mathscr{R} - \mathscr{H} * (\mathscr{T} - \mathscr{L}_t) - \frac{\mathfrak{Y}_{t-1}}{\beta}\right\|_F^2.\end{aligned} \tag{19}$$

The problem of equation (19) can be used as the proximal operator of the ℓ_1 norm to solve

$$\mathscr{R}_t = \mathrm{Prox}_{\frac{\lambda}{\beta}\|\cdot\|_1}\left(\mathscr{H} * (\mathscr{T} - \mathscr{L}_t) + \frac{\mathfrak{Y}_{t-1}}{\beta}\right). \tag{20}$$

Equation (20) can be efficiently computed by the element-wise soft thresholding operation and is as follows:

$$\mathrm{Prox}_{\frac{\lambda}{\beta}\|\cdot\|_1}\left(\mathscr{F}_{i.j.k}\right) = \mathrm{sign}\left(\mathscr{F}_{i.j.k}\right)\max\left\{\left|\mathscr{F}_{i.j.k}\right| - \frac{\lambda}{\beta}, 0\right\}, \tag{21}$$

where

(3) Update $\mathfrak{Y}_t$ as follows:

$$\mathfrak{Y}_t = \mathfrak{Y}_{t-1} + \beta(\mathscr{H} * (\mathscr{T} - \mathscr{L}_t) - \mathscr{R}_t). \tag{22}$$

(4) *Stopping Criterion*. Given a tolerance $\varepsilon > 0$, we check the termination condition of primal variables $\|\mathscr{L}_t - \mathscr{L}_{t-1}\|/\|\mathscr{T}\| \le \varepsilon$.

After discussing the appearing subproblem, the complete DR-SKSNM algorithm for sensor data reconstruction in NB-IoT is presented in Algorithm 1.

The computational complexity of DR-SKSNM consists mainly of two parts: one is the complexity of the relative density-based clustering algorithm and the other is the cost of tensor reconstruction computation. When the relative density-based clustering algorithm is used to separate sensor nodes, c cluster centers are first set randomly, and then computing the distance from n points to c centers, the distance between the single node and cluster center is calculated as P, and finally the above process is repeated for t times, so the complexity of relative density-based clustering is O(cnpt). Usually, c, P, and t can be regarded as constants; therefore, the computational cost of relative density-based clustering can be simplified to be linear, namely, O(n). From Algorithm 1, the sensor data reconstruction algorithm mainly involves computation of $\mathscr{L}_t$,$\mathscr{R}_t$, and $\mathfrak{Y}_t$, the parameters iterated T times, so the computational complexity of the matrix completion algorithm is O(nT). Thus, the computational complexity of DR-SKSNM in this paper is O(nT).

4. Experimental Results

In this section, the performance of our proposed algorithm called DR-WSKSNM is investigated by comparing it over other alternative benchmark algorithms.

4.1. Data Description. The water quality monitoring NB-IoT platform in Sichuan Province (China) from industrial and urban sewage discharge monitoring stations was utilized to collect original data. The monitoring variables including chemical oxygen demand (COD) (unit:mg/L), ammonia nitrogen (unit:mg/L), total nitrogen (unit:mg/L), and pH are measured and stored in Sichuan Province intelligent ecological management platform, every 10 minutes from January 1, 2020, to the present. The data from January 1, 2020 (00 : 00 : 00 a.m.) to August 30, 2020 (23 : 59 : 59 p.m.) are particularly extracted to utilize for analyzing reconstruction performance. Table 1 shows some analyzing samples from sensor ID 105 of the water quality monitoring NB-IoT platform.

To validate the availability of the algorithm objectively, two types such as random missing pattern and consecutive missing pattern are applied to verify [11].

(i) *Random Missing Pattern (RMP)*. This pattern is suitable for missing data with a random time and random sensor to be missing. α represents the data missing rate.

(ii) *Sequence Missing Pattern (SMP)*. This pattern reflects that all data are missed after a certain sampling time point owing to running out of batteries or experiencing a loss of connectivity to the acquisition platform. Therefore, we randomly selected 5% of nodes as objective nodes that suffer from sequence data missing. Then, the missing subsequences with different time intervals such as one-hour, three-hour, six-hour, and nine-hour missing are utilized to assess the validity of the proposed method.

Input: T, H;
Initialize: $L_0 = T, L_1 = 0, R_0 = 0, S_0 = 0$ and set parameters λ, β, δ, ε, $t = 1 \ldots T$.
While $\|L_t - L_{t-1}\|/\|T\| \geq \varepsilon$, then
  $t = t + 1$;
  Update L_t by equation (17);
  Update R_t by equation (20);
  Update S_t by equation (21);
End while
Output: L_{t-1}.

Algorithm 1: DR-SKSNM algorithm for NB-IoT sensor data reconstruction.

Table 1: Data samples from the water quality monitoring NB-IoT platform.

Monitoring time	COD (mg/L)	Ammonia nitrogen (mg/L)	Total nitrogen (mg/L)	pH
2020-01-01 00 : 00 : 00	12.009	2.419	9.21	7.02
2020-01-01 00 : 10 : 00	12.27	2.124	9.27	7
2020-01-01 00 : 20 : 00	12.239	2.07	9.28	6.98
2020-01-01 00 : 30 : 00	11.81	1.923	9.41	6.95
...	...	...	...	...

4.2. Reconstruction Performance Indicators. Our algorithm is evaluated with the existing data reconstruction methods based on the following two metrics.

4.2.1. MAPE (Mean Absolute Percentage Error). MAPE (p, q) is defined as the average of absolute percentage errors of forecasts, which is measured by the following:

$$\mathrm{MAPE}(p, q) = \frac{100}{w} \sum_{i=1}^{w} \left| \frac{p_i - q_i}{q_i} \right|, \tag{23}$$

where w denotes the length of missing subsequence, P is a predicted value, and q indicates a true value. A smaller $\mathrm{MAPE}(p, q)$ signifies a better approach for the reconstruction of missing values.

4.2.2. RMSE (Root Mean Square Error). RMSE (P, q) is a metric for measuring reconstruction error and is denoted as the average squared difference between the predicted data (P) and the true data (q). A better performance method will have a lower value of RMSE. It is defined as follows:

$$\mathrm{RMSE}(p, q) = \sqrt{\frac{1}{w} \sum_{i=1}^{w} (p_i - q_i)^2}. \tag{24}$$

In our experiment, we set some parameters empirically [20], specifically, $\lambda = 0.4$, $\beta = 0.5$, $\delta = 0.01$, and $\varepsilon = 10^{-3}$. The parameter ε affects the speed of convergence of the ADMM algorithm. All simulations were run in the MATLAB 2016b environment on a laptop equipped with an Apple M1 processor and 8 GB RAM.

4.3. Performance Comparison and Discussion. To examine the performance of our proposed algorithm, four different methods based on tensor data reconstruction are compared with the true values for validation. These true values are set as ground truth to make quantitative comparisons by utilizing two metrics, namely, MAPE and RMSE.

Figure 3 shows the locations of sensor nodes and their clusters in the Panzhihua region from China. Here, the number of clusters is 5. The corresponding sensor nodes of each cluster are included within the red solid lines in Figure 3. Our DR-SKSNM algorithm for missing data reconstruction was applied within each group except cluster 5 due to fewer sensor nodes in it. Tensors are composed of attributes of monitor index, nodes, and time slots in each cluster. In our experiment, the outputs of the recovered sensor data with that of existing algorithms, namely, DRAWNNM [16], HaLRTC [20], and ADMAR [21] were compared with the output of the proposed method. By comparison, the correct rank (3, 2, 3) and a higher rank are set to execute Tucker decomposition in random missing patterns and consecutive missing patterns.

Tables 2 and 3 demonstrate, respectively, the quantitative results of four methods on indicators for the task of reconstructing missing values with missing rate of RMP and missing gap of SMP. The best indicator values are labeled as italic and bold in Tables 2 and 3. From the comparison, the performance of our algorithm is better than that of the existing algorithms from two metrics on the COD and total nitrogen datasets. When the missing rate is less than 20%, the DRAWNNM algorithm performs as well as the other methods. Moreover, when the missing rate is higher than 20%, our indicator values are the best. When comparing the performance of DR-SKSNM on two different datasets, better performance in terms of MAPE and RMSE appears on the COD dataset than on the total nitrogen dataset. This is because the COD dataset has preferable monitoring data that are more similar among measured values of neighboring sensors. The result of total nitrogen dataset is particularly unstable because it is based on the influence of parameter drifting on sampling data.

Table 2: Performance evaluation for four methods by average MAPE and RMSE on the COD and total nitrogen datasets (unit: mg/L) with RMP.

Dataset	Missing rate	DRAWNNM		HaLRTC		ADMAR		Proposed method	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
COD	$\alpha = 10\%$	15.745	38.475	16.701	39.421	16.945	39.584	14.546	38.491
	$\alpha = 20\%$	42.608	85.270	43.624	88.130	43.123	94.220	40.964	85.560
	$\alpha = 40\%$	99.641	189.678	91.248	195.781	90.478	198.657	88.735	188.953
Total nitrogen	$\alpha = 10\%$	6.175	45.285	7.930	48.698	7.270	52.357	6.190	45.025
	$\alpha = 20\%$	27.457	89.324	28.907	93.234	29.098	94.876	26.896	87.324
	$\alpha = 40\%$	86.323	175.239	88.680	168.853	89.510	158.680	80.764	155.381

Table 3: Performance evaluation for four methods by average MAPE and RMSE on the COD and total nitrogen datasets (unit: mg/L) with SMP.

Data type	Missing gap	DRAWNNM		HaLRTC		ADMAR		Proposed method	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
COD	One-hour	15.704	27.360	17.693	26.030	16.942	19.660	15.067	23.140
	Three-hour	32.439	93.060	33.641	83.780	32.380	92.730	30.290	70.570
	Six-hour	72.110	139.570	89.820	141.330	99.550	147.910	64.980	137.640
	Nine-hour	131.983	265.522	126.638	253.591	131.628	280.032	119.190	245.730
Total nitrogen	One-hour	23.589	29.452	25.159	32.021	24.929	37.230	23.976	33.381
	Three-hour	45.178	96.271	48.627	93.737	46.680	94.560	40.221	89.321
	Six-hour	82.275	169.601	86.921	163.260	90.172	177.290	84.217	157.290
	Nine-hour	156.210	274.289	147.210	283.240	158.258	288.126	149.274	265.216

The corresponding results for each dataset are illustrated in Figures 4–7. First, on the COD dataset, Figure 4 demonstrates the visual comparison of reconstruction value curves of four methods and real values curve at position node 26 in cluster 4 from Figure 3 with a missing gap of six-hour. As noticed in Figure 4, the curve shape of reconstructed values resulted from DR-SKSNM only corresponds more closely to that of real values. This substantiates that DR-SKSNM yields superior performance than other compared methods when sensors in one group have similar patterns of measurement. The reconstructed values from DRAWNNM, HaLRTC, and ADMAR also match less with real values than DR-SKSNM in Figures 4–6 though they do generate the up-down trends.

As illustrated in Figures 5 and 6, our algorithm merely demonstrates its ability with missing gaps of nine-hour and twelve-hour for ammonia nitrogen and total nitrogen datasets at position node 36 in cluster 5 and node 21 in cluster 3 from Figure 3, respectively. Our proposed method has better consistency between reconstructed data and real data from Figures 5 and 6. Because the reconstructed data curves of compared methods fluctuate greatly, there are many deviations from the real data. In particular, it is not difficult to find the curve trend from Figure 6 that the local error of reconstructed data of the ADMAR method is high. Notably, when the data missing gaps are high, the error rate of the proposed algorithm is considerably lower than that of the other three algorithms. The reason is that our algorithm DR-SKSNM uses the relative density-based clustering algorithm to group sensors, which greatly enhances the connection of sensor data and improves the accuracy of the algorithm. As shown in Figure 7, a visual comparison of the

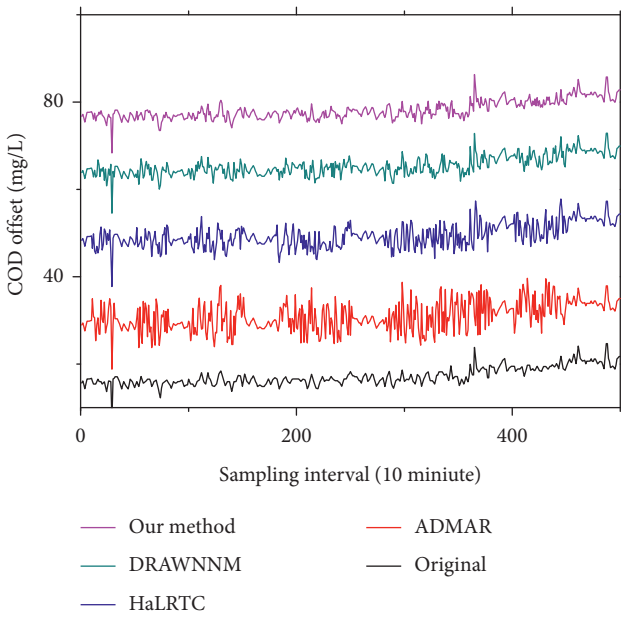


Figure 4: Visual comparison of reconstruction curves of four methods and original data curve (monitoring variable dataset: COD).

reconstruction curves of four approaches and the original data curve for pH datasets with missing gaps of two hours at location node 26 from Figure 3 is reported. From Figure 7, it can be seen that the DR-SKSNM algorithm also has a good performance and can reconstruct the missing data well. Moreover, the DR-SKSNM algorithm is superior to the other algorithms, including DRAWNNM.

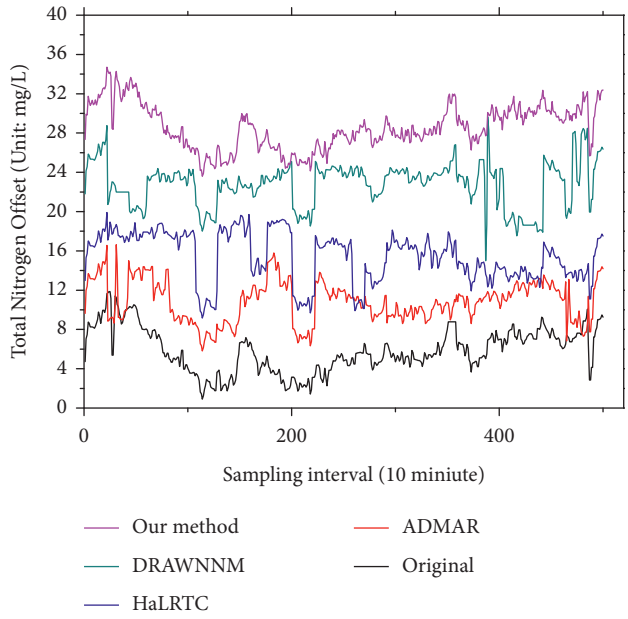


Figure 5: Visual comparison of reconstruction curves of total four methods and original data curve (monitoring variable dataset: total nitrogen).

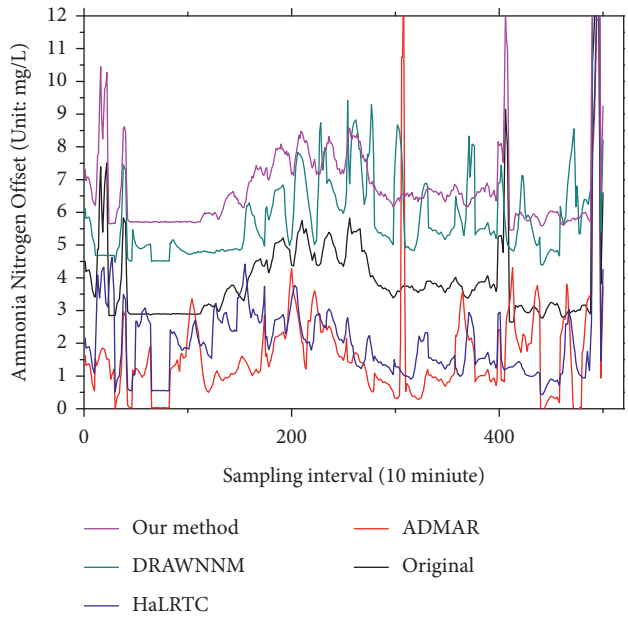


Figure 6: Visual comparison of reconstruction curves of total four methods and original data curve (monitoring variable dataset: ammonia nitrogen).

Consequently, the reconstruction performances of all the existing methods including the proposed algorithm decisively vary on the characteristics of each dataset for testing. Thus, the existing techniques are not as effective as our proposed algorithm, and we further confirm that DR-SKSNM is a better option for reconstruction missing data in real-life NB-IoT applications.

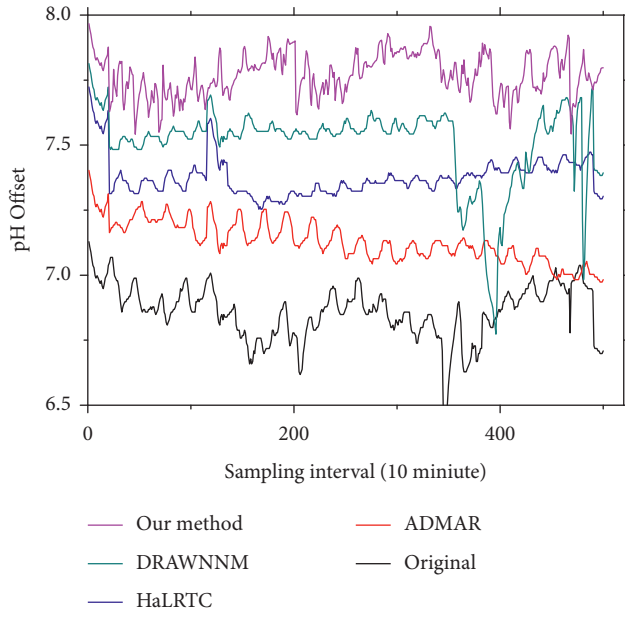


Figure 7: Visual comparison of reconstruction curves of four methods and original data curve (monitoring variable dataset: pH).

5. Conclusion

Missing data from sensors hinder many NB-IoT applications. To solve the problem, it is necessary to estimate the missing data as accurately as possible. In this paper, we presented the DRAWNNM algorithm as a novel approach to reconstruct missing data for NB-IoT applications. This algorithm has been tested using four different real datasets currently stored in our databases. We examined the accuracy of reconstructed values generated from DR-SKSNM, by comparing with three existing methods such as DRAWNNM, HaLRTC, and ADMAR, using two performance metrics (MAPE and RMSE). Through the experiments, we demonstrate that our proposed algorithm outperforms these existing methods when dealing with high data missing rates and large data missing gaps as validated both by performance in terms of MAPE and RMSE. Overall, this work provides strong evidence that DRAWNNM is potentially superior to compared algorithms in terms of accuracy and computational complexity. Finally, the intriguing future work in this paper is that the correlation among more multiple various sensors is taken into account and machine learning techniques are used to identify a superior subsequence.

Data Availability

The data used to support the findings of this study were supplied by luoxuegang under license and so cannot be made freely available. Requests for access to these data should be made to corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Innovation Foundation of Sichuan Province of China under Grant 2019088, and in part by the Science and Technology Support Project of Panzhihua City of Sichuan Province of China under Grant 2021CY-S-6.

References

[1] M. Ganzha, M. Paprzycki, and W. Pawlowski, "Semantic interoperability in the Internet of Things; an overview from the INTER-IoT perspective," *Journal of Network and Computer Applications*, vol. 81, no. MAR., pp. 111–124, 2016.

[2] M. V. Ramesh, K. V. Nibi, and A. Kurup, "Water quality monitoring and waste management using IoT," in *Proceedings of the 2017 IEEE Global Humanitarian Technology Conference (GHTC)*, IEEE, San Jose, CA, USA, October 2017.

[3] L. Teng, X. Min, and J. Chen, "Automated water quality survey and evaluation using an IoT platform with mobile sensor nodes," *Sensors*, vol. 17, no. 8, p. 1735, 2017.

[4] S. Shanzhi Chen, H. Hui Xu, D. Dake Liu, B. Hu, and H. Wang, "A vision of IoT: applications, challenges, and opportunities with China perspective," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 349–359, 2014.

[5] J. Lou and Y.-M. Cheung, "Robust low-rank tensor minimization via a new tensor spectral k -support norm," *IEEE Transactions on Image Processing*, vol. 29, no. 1, pp. 2314–2327, 2020.

[6] Y. Wang and Y. Yang, "Relative density-based clustering algorithm for identifying diverse density clusters effectively," *Neural Computing & Applications*, vol. 33, no. 16, pp. 10141–10157, 2021.

[7] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TinyDB: an acquisitional query processing system for sensor networks," *ACM Transactions on Database Systems*, vol. 30, no. 1, pp. 122–173, 2005.

[8] L. Pan and J. Li, "K-nearest neighbor based missing data estimation algorithm in wireless sensor networks," *Wireless Sensor Network*, vol. 2, no. 2, pp. 115–122, 2010.

[9] S. A. Rahman, Y. Huang, J. Claassen, N. Heintzman, and S. Kleinberg, "Combining Fourier and lagged k -nearest neighbor imputation for biomedical time series data," *Journal of Biomedical Informatics*, vol. 58, pp. 198–207, 2015.

[10] Y. Li and L. E. Parker, "Nearest neighbor imputation using spatial-temporal correlations in wireless sensor networks," *Information Fusion*, vol. 15, pp. 64–79, 2014.

[11] G. Ho Lee, J. Han, and J. K. Choi, "MPdist-based missing data imputation for supporting big data analyses in IoT-based applications," *Future Generation Computer Systems*, vol. 125, pp. 421–432, 2021.

[12] Y. Zaid, B. Zhang, W. M. Ismael, Y. Xie, G. N. Surname, and H. Wang, "A spatio-temporal multiple linear regression missing data reconstruction approach for improving wsn data reliability," in *Proceedings of the 2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, pp. 1–6, July 2021.

[13] J. Zhang and P. Yin, "Multivariate time series missing data imputation using recurrent denoising autoencoder," in *Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 760–764, IEEE, San Diego, CA, USA, November 2019.

[14] P. Thi-Thu-Hong, "Machine learning for univariate time series imputation," in *Proceedings of the 2020 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, IEEE, Hanoi, Viet Nam, October 2020.

[15] N. Bokde, M. W. Beck, F. Martínez Álvarez, and K. Kulat, "A novel imputation methodology for time series based on pattern sequence forecasting," *Pattern Recognition Letters*, vol. 116, pp. 88–96, 2018.

[16] X. Yu, X. Fan, K. Chen, and S. Duan, "Multi-attribute missing data reconstruction based on adaptive weighted nuclear norm minimization in IoT," *IEEE Access*, vol. 6, pp. 61419–61431, 2018.

[17] A. K. Jain, "Data clustering: 50 Years beyond K-means," in *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2008*, W. Daelemans, B. Goethals, and K. Morik, Eds., vol. 5211, Springer, Berlin, Heidelberg, 2008, Lecture Notes in Computer Science.

[18] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[19] A. M. Mcdonald, M. Pontil, and D. Stamos, "Fitting spectral decay with the k-support norm," in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (PMLR)*, vol. 51, pp. 1061–1069, Cadiz, Spain, May 2016.

[20] Y. Shao, Z. Chen, F. Li, and C. Fu, "Reconstruction of big sensor data," in *Proceedings of the 2nd IEEE International Conference Computer Communication (ICCC)*, pp. 1–6, Chengdu, China, October 2016.

[21] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 4170064, 15 pages
https://doi.org/10.1155/2021/4170064

Research Article

Joint Matrix Decomposition-Based Missing Data Completion in Low-Voltage Area

Haowen Wu,[1] Chen Yang,[1] Wenwang Xie,[1] and Wei Zhang[2]

[1]*China Southern Power Grid Digital Grid Research Institute Co., Ltd., Guangzhou, Guangdong 510663, China*
[2]*Hefei University of Technology, Hefei, Anhui 230009, China*

Correspondence should be addressed to Wei Zhang; 2020170417@mail.hfut.edu.cn

Received 23 August 2021; Revised 30 September 2021; Accepted 18 October 2021; Published 8 November 2021

Academic Editor: Xianyong Li

In-depth mining and analysis of electricity data in low-voltage area are essential for the further intelligent development of power grids. However, in the actual data collection and measurement of low-voltage area, there will be missing data, and complete electricity data cannot be obtained. To obtain complete power data, this paper proposes a low-voltage station area missing data complement model based on joint matrix decomposition. First, we analyse the characteristics of the low-pressure station data. Then, a model that comprehensively considers the characteristics of the low-voltage station area data is proposed, which includes three parts: the construction of a low-voltage station area data tensor, the joint matrix decomposition, and the completion of the missing data, and it is named LPZ. After that, the CIM learning algorithm proposed in this paper is used to iteratively solve the model to obtain the completed data. Finally, the method proposed in this paper is used to complement the two situations of random loss and all-day loss of real current data in a low-voltage station area and compared with the traditional complement method. The experimental results show that this method is not only effective but also that the completion effect is better than that of other completion methods.

1. Introduction

In recent years, with the continuous advancement of intelligent power grid construction, in-depth mining and analysis of electricity data have become increasingly important [1, 2]. Electricity data contain a large amount of electricity consumption data information. Through in-depth mining and analysis of electricity consumption data, various advanced applications, such as electricity demand and electricity price setting, can be realized to provide support for the safe and efficient operation of the power grid [3–5]. As an important part of the power grid, the low-voltage station area, in-depth mining, and analysis of its electricity data will become the key to further intelligentization of the power grid, which has important significance for the future [6].

To successfully realize the in-depth mining and analysis of the electricity data of the low-voltage station area, it is necessary to maintain the integrity of its electricity consumption data as much as possible. In the actual process of the low-voltage station area, the data are missing due to equipment damage, weather conditions and other reason, causing a sharp fall in the quality. For example, the data in the literature [7] have a missing rate of multiple attributes exceeding 50%, and there are missing data almost in every record. Therefore, in the in-depth mining and analysis of electricity data, it is necessary to process the missing data with the known data [8, 9].

At present, scholars at home and abroad have carried out some researches on the methods of completing missing data. Literature [10] attempted to complement the synchrophasor measurement data of the synchrophasor measuring device using the matrix filling method. However, as low-voltage station area power data have different characteristics from synchronized phasor measurement data, that is, the similarity of the power consumption data of each user is quite different, it is difficult to directly apply the matrix filling theory to achieve the repair effect. Literature [11] utilized an

adaptive neuro-fuzzy inference system model to complement and optimize the missing power data based on the traditional single data completion method (such as interpolation). The aforementioned methods require a large amount of data for pretraining. Thus, a small sample size may lead to unsatisfactory effect of completion. Literature [12] uses the KNN completion algorithm to complete missing values. Although the KNN completion algorithm is simple, intuitive, and easy to implement without the need for prior knowledge, the accuracy of its completion depends on the average value of the neighboring sample data. Literature [13] based on the single-value missing data completion method of nonlocal averaging method proposed a multivalue missing data completion method using spatial neighbor BP (backpropagation) mapping to achieve higher precision data completion. However, this method does not consider the timing of the data to be completed. Literature [14] considered that in the actual system, a subset of the data may have stronger relevance, and proposed a new local tensor completion model. This model uses the stronger local correlation of the data to form and restore each subtensor with a lower level to achieve accurate data recovery. However, this method cannot achieve data completion when there is less correlation between the data. Literature [15] adopted the machine learning-based method to perform missing value completion. Although this type of method performs well in accuracy, it is necessary to learn the complete data sequence, and it is difficult to find the complete sequence training parameters in actual search.

Given the above problems, this paper proposes a low-voltage station data completion method based on joint matrix decomposition based on the characteristics of the low-voltage station area electricity data. First, we analyse the characteristics of the low-pressure station data. Then, a model that comprehensively considers the characteristics of the low-voltage station area data is proposed. The modified model includes three parts: the construction of a low-voltage station area data tensor, joint matrix decomposition, and completion of the missing data, and it is named LPZ. After that, the CIM learning algorithm proposed in this paper is used to iteratively solve the model to obtain the completed data. Finally, through the verification of real current data of a low-voltage station area, the effectiveness of the method proposed in this paper is obtained, and compared with the traditional complement method, the superiority of the method in this paper is obtained.

2. Analysis of the Characteristics of Electricity Data in Low-Voltage Area

The low-voltage station area electricity data mainly include the voltage, current, active power, reactive power, and other data of each user in the station area [16]. The data mainly have the following characteristics:

(1) *Periodicity*. Generally, the power data of low-voltage area shows periodic changes over several consecutive working days; that is, on consecutive working days, the electricity consumption behavior of each user shows a similar periodic law. As shown in Figure 1, on different working days, the current curve of a user in the station area has a similar trend.

(2) *Sequentiality*. User data all appear in the form of data streams, which are sequentially collected, transmitted, and stored at equal time intervals. The data analysed in this paper are based on a sampling interval of 30 minutes, and 48 points of data are collected a day. The data graph with the sampling interval as the time window is shown in Figure 2.

(3) *Spatial Correlation*. In the power system, different users are connected through the network topology of the station area, and the power load between different users has a certain correlation, especially when a high-power electrical appliance starts or malfunctions. The performance will be more obvious. Therefore, it is necessary to consider the multiuser spatial correlation of the station area data to complete the missing data of multiple users.

3. Complementary Method for Missing Data in Low-Voltage Station Areas Based on Joint Matrix Decomposition

3.1. Model Structure. According to the analysis of the characteristics of low-voltage station area electricity data, we conclude that the low-voltage station area data contain three characteristics: periodicity, time series, and spatial correlation. Therefore, we propose a model that comprehensively considers the data characteristics of the low-voltage station area to solve the problem of data completion in the low-voltage station area and name it LPZ. Specifically, first, we design the organization of electricity data. The electricity consumption data sequence of all users in the low-voltage station area is organized into a tensor, which can not only mine potentially related information from different patterns but also ensure the original characteristics of the station area electricity data. Next, we propose a joint decomposition module [17] that extracts the day-time interval matrix and the user-time interval matrix from this tensor and then decomposes all the extracted matrices, that is, an original matrix is expressed in the form of the product of two low-dimensional matrices to form an expression of all users, days, and time intervals. Furthermore, to mine the characteristics of the temporality of the data, we added a local restriction to the joint decomposition module. The restriction condition we adopted here is to make the predicted value of the target value close to the predicted value of the adjacent time interval.

The LPZ architecture is shown in Figure 3, including three parts: construction of a low-voltage station area data tensor, joint matrix decomposition, and missing data completion. Among them, the two decomposition modules in the joint decomposition matrix provide the characteristic expression of users, days, and time intervals by mining potential factors, and both decomposition modules are affected by local restrictions, which makes the model consider the period of low-voltage station data. It is also possible to

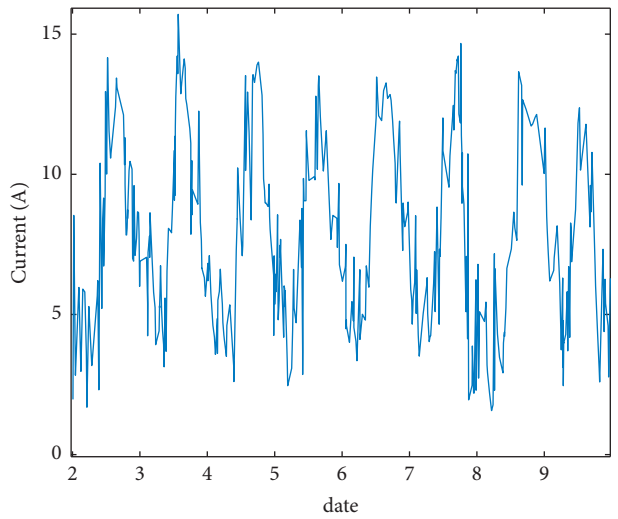


Figure 1: Current curve of a user for 8 consecutive days.

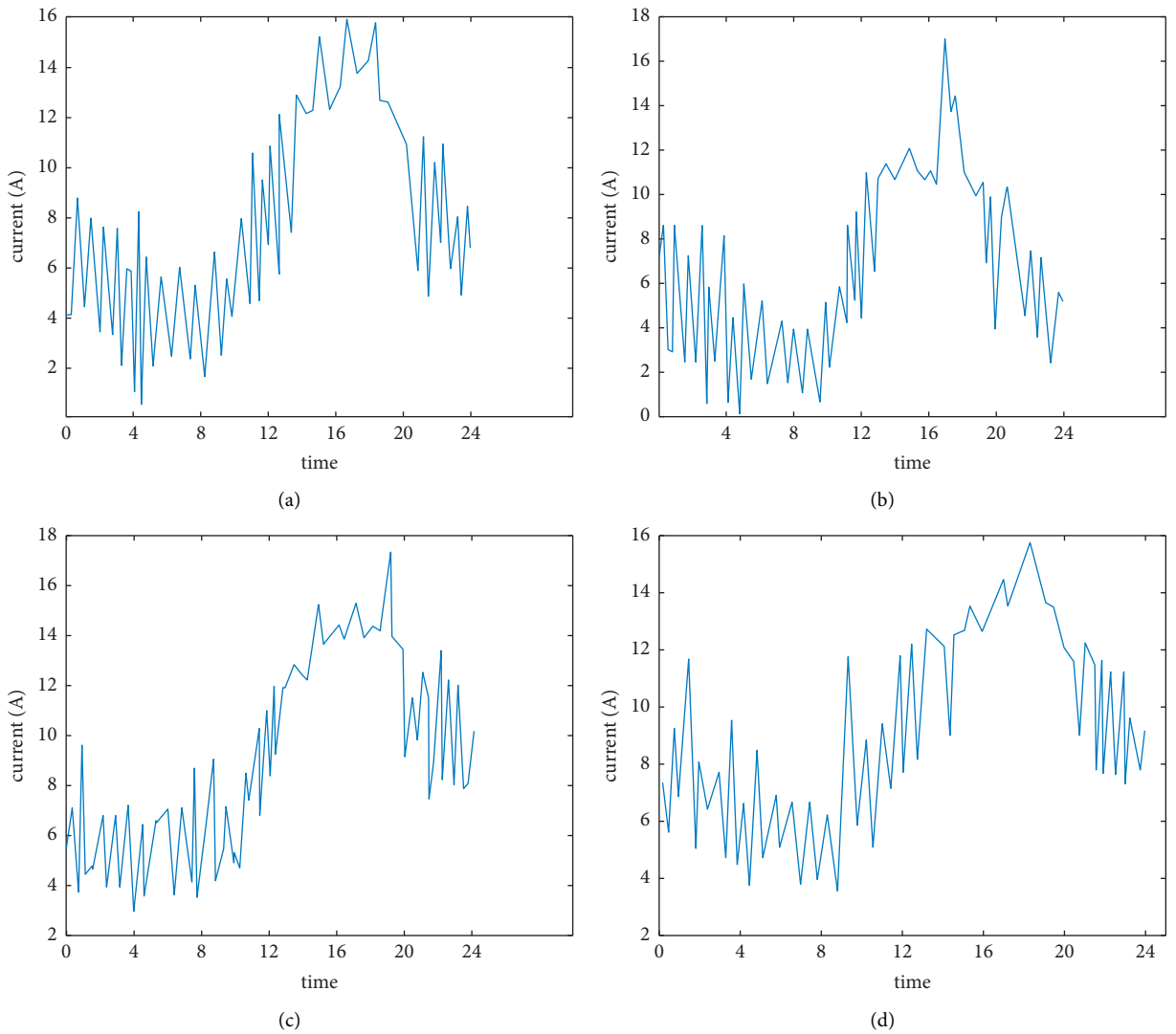


Figure 2: The current size of the same user on four days in a month. (a) First day's data. (b) Fourth day's data. (c) Eighth day's data. (d) Twelfth day's data.

mine and utilize the temporality of data as well as spatial and spatial relevance.

3.2. Constructing the Data Tensor of the Low-Pressure Station Area. Based on the characteristics of low-voltage station area electricity data, we designed the input of the model. First, we assume that the power consumption data of user i in time interval j on day k is $c_{i,j,k}$; then, the power consumption data sequence obtained by user i in chronological order is $S_i = <c_{i,0,0}, \ldots, c_{i,j,0}, c_{i,0,1}, \ldots, c_{i,J,1}, \ldots, c_{i,J,K}>$, where J is the number of time intervals in a day and K represents the number of days. We fold it into K vectors, and each vector contains the electricity consumption data of the user at various time intervals in a day. Then, these vectors are integrated to form a two-dimensional matrix form, as shown in equation:

$$\begin{bmatrix} c_{i,0,0} & c_{i,1,0} & \cdots & c_{i,J,0} \\ c_{i,0,1} & c_{i,1,1} & \cdots & c_{i,j,1} \\ \vdots & \vdots & \ddots & \vdots \\ c_{i,0,K} & c_{i,1,K} & \cdots & c_{i,J,K} \end{bmatrix} \in R^{K\times J}. \tag{1}$$

Performing a folding operation on the electricity consumption data sequence of I users can obtain I two-dimensional matrices. Then, we integrate them into a three-dimensional tensor (user, time interval, and day) and use it as input to the model. This input not only maintains the characteristics of the original data but also makes the tensor pattern closely related to the characteristics of the low-voltage station area electricity data.

3.3. Joint Matrix Factorization. Due to the regularity of people's activities, on a continuous working day, a user's electricity consumption data will generally show periodic changes. To model the periodicity of user electricity data, we extract the number of day-time interval matrix $\mathbf{C}_i$ from the electricity data tensor C and learn the temporality from $\mathbf{C}_i$ through matrix decomposition. Then, hidden factors are introduced, and matrix $\mathbf{C}_i$ is decomposed into two low-dimensional matrices $\mathbf{P}_i \in R^{K\times F}$ and $Q_i \in R^{J\times F}$, where F is the number of hidden factors. In addition, each day k is related to a vector $p_k \in R^F$. Similarly, each time interval j corresponds to a vector $q_j \in R^F$, where the similarity between the electricity consumption data of different days or different time intervals can be captured in a potentially low-dimensional space. Thus, the predicted value $c_{i,j,k}$ of the power consumption data $\widetilde{c}_{i,j,k}$ of user i in time interval j of day k is obtained, which is represented by the inner product $p_k \cdot q_j$.

At the same time, the electricity consumption data series of different users will also be correlated. Therefore, we also decompose the user-time interval matrix $\mathbf{C}_k$ to explore the spatial correlation of electricity consumption data. All users and time intervals are mapped to a low-dimensional space. In this latent space, $\mathbf{S}_k \in R^{I\times F}$ represents all users, and $\mathbf{T}_k \in R^{J\times F}$ represents all time intervals. Moreover, each user i corresponds to a vector $s_i \in R^F$, and each time interval j corresponds to a vector $t_j \in R^F$. Thus, the predicted value $\widetilde{c}_{i,j,k}$ of the power consumption data $c_{i,j,k}$ of user i in time interval j of day k is obtained, which is expressed by the inner product $s_i \cdot t_j$. To obtain more accurate prediction results, when predicting the missing low-voltage station data, we combine the two matrix decomposition modules to take periodicity and spatial correlation into account at the same time.

In addition, the electricity consumption data of a certain time interval have a strong correlation with the electricity consumption data of the surrounding time interval. Therefore, we introduce local constraints into the joint matrix factorization process. In the process of decomposing the number of day-time intervals, we also need to minimize the difference between the predicted value of the target electricity consumption data and the mean value of the surrounding time interval, as shown in the following equation:

$$\begin{aligned} g_1 &= \left(p_k \cdot q_j - \overline{c}^{(1)}_{i,j,k}\right)^2, \\ \overline{c}^{(1)}_{i,j,k} &= \frac{1}{2W}\sum_{\omega}^{W}\left(p_k \cdot q_{j-\omega} + p_k \cdot q_{j-\omega}\right), \end{aligned} \tag{2}$$

where W is the window size, and $\omega = 1, 2, \ldots, W$.

Similarly, a local restriction is also added in the decomposition process of the user-time interval matrix, as shown in following equation:

$$\begin{aligned} g_2 &= \left(s_i \cdot t_j - \overline{c}^{(2)}_{i,j,k}\right)^2, \\ \overrightarrow{c}^{(2)}_{i,j,k} &= \frac{1}{2W}\sum_{\omega=1}^{W}\left(s_i \cdot t_{j-\omega} + s_i \cdot t_{j+\omega}\right), \end{aligned} \tag{3}$$

where W is the window size, and $\omega = 1, 2, \ldots, W$.

In summary, the low-voltage station area data contain three characteristics, namely, periodicity, time series, and spatial correlation. In addition, the collected data contain a large number of missing values, causing data sparsity problems. Therefore, we design and use the joint matrix decomposition module to model the periodicity and spatial correlation, respectively. At the same time, we set local constraints based on the spatial correlation to restrict the joint decomposition module. In this way, the three features can work synergistically when completing missing values.

3.4. Completion of Missing Data. Through joint matrix decomposition, we can obtain the hidden factor matrices $\mathbf{P}_i$ and $\mathbf{Q}_i$ of user i and the hidden factor matrices $\mathbf{S}_k$ and $\mathbf{T}_k$ of the number of days k. Finally, we obtain four-parameter tensors, namely, $P \in R^{K\times F\times I}$, $\mathbf{Q} \in R^{J\times F\times I}$, $\mathbf{S} \in R^{I\times F\times K}$, and $\mathbf{T} \in R^{J\times F\times K}$. Using these factor tensors can realize the completion of the original incomplete matrix. Here, we use a simple regression method to combine the two partial results and use it as the output of the joint matrix factorization module. The weight β is set to control this combination process, where the weight β represents the periodic force, and $(1-\beta)$ represents the weight of the spatial correlation in

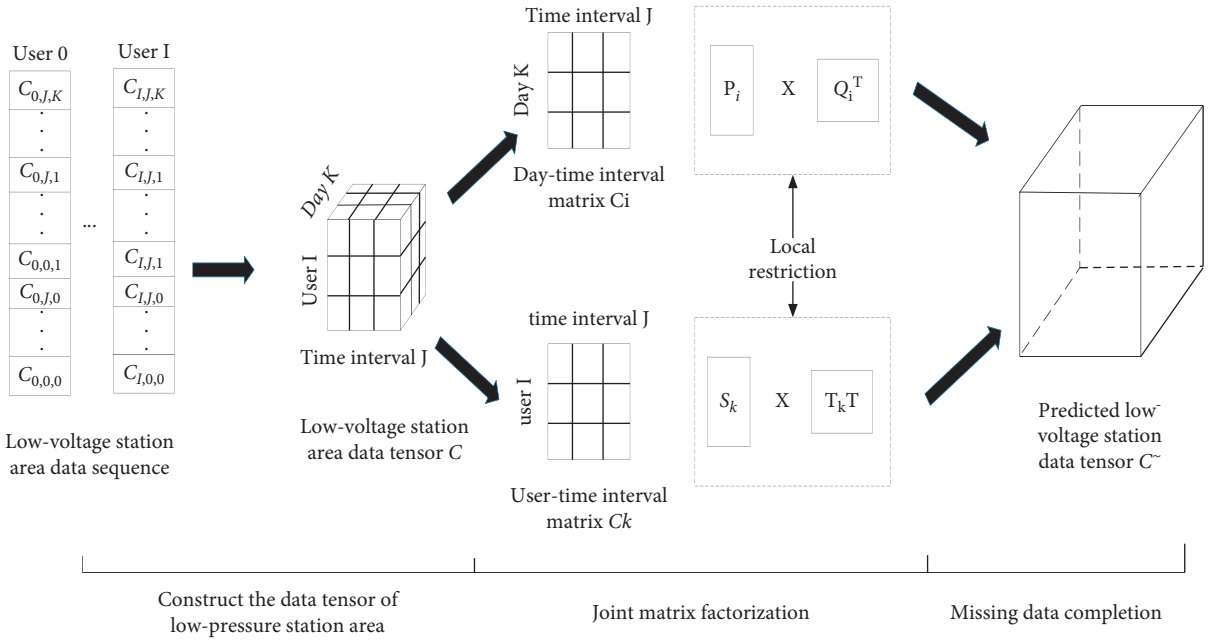


Figure 3: LPZ architecture diagram.

the missing value prediction process. This linear fitting process enables LPZ to reasonably weigh the characteristics of the low-voltage station area data, thereby obtaining a complete low-voltage station area electricity data tensor. Thus, the prediction formula of the missing value shown in formula (4) is obtained:

$$\widetilde{c}_{i,j,k} = \beta P_{k,:,i} Q^T_{j,:,i} + (1-\beta) S_{i,:,k} T^T_{j,:,k}, \tag{4}$$

where $\mathbf{P}_{k,:,i} \in \mathbf{R}^F$ is equivalent to p_k; $\mathbf{Q}_{j,:,i} \in R^F$ is equivalent to q_j; $\mathbf{S}_{i,:,k} \in R^F$ and s_i are equivalent; and $\mathbf{T}_{j,:,k} \in R^F$ is equivalent to t_j.

3.5. Objective Function and Parameter Learning. We establish the model objective function by minimizing the square difference between the true value of the low-voltage station area electricity data and its estimated value. Given a low-voltage station area electricity data tensor C, this tensor contains a large number of missing values. Our model can mine multiple features of the low-voltage station area electricity data and generate one and the original tensor based on the known values in the tensor, completely estimating tensors with the same shape to realize the task of complementing low-pressure station area data.

For simplicity, we set a binary mask $\mathbf{M}$; this mask tensor corresponds to the original tensor, and its value is also determined by the value of the element at the corresponding position in the original tensor. In the masking tensor, the missing element in the original low-voltage station area electricity data tensor has a value of 1, and the observed element position is 0; that is, through the masking tensor $\mathbf{M}$, we can clearly know the missing values in the original tensor, and the position of the observation value is formulated as the following equation:

$$m_{i,j,k} = \begin{cases} 1, & c_{i,j,k} \text{ missing}, \\ 0, & c_{i,j,k} \text{ known}. \end{cases} \tag{5}$$

Therefore, the missing value in the original matrix can be expressed as $\mathbf{M} \odot \mathbf{C}$, and the observed value can be expressed as $(1 - \mathbf{M}) \odot \mathbf{C}$. The objective function can be defined as the following formula:

$$\ell = \|(1-\mathbf{M}) \odot (\widetilde{C} - \mathbf{C})\|_F^2 + \lambda \|\theta\|^2, \tag{6}$$

where $\odot$ represents the point multiplication operation, $\|\cdot\|_F$ represents the Frobenius norm, $\lambda\|\theta\|^2$ represents the regularization term to prevent overfitting, and θ represents all parameter tensors.

According to the description in the previous two subsections, to consider the periodicity and spatial correlation of the low-voltage station area data at the same time, we designed a joint decomposition module and added local restrictive constraints in the decomposition process so that LPZ can consider the electricity data at the same time. Here, we use the linear fitting method to combine the local results of the two decomposition models to obtain the objective function shown in the following equations:

$$
\begin{aligned}
l = & \underbrace{\beta \sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{k=0}^{K}\left(\mathbf{P}_{k,:,i}Q_{j,:,i}^{T}-c_{i,j,k}\right)^{2}}_{Spatial\ correlation} \\
& + \underbrace{(1-\beta)\sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{k=0}^{K}\left(S_{i,:k}T_{j,:,k}^{T}-c_{i,j,k}\right)^{2}}_{Periodic} \\
& + \underbrace{\lambda_1 \sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{k=0}^{K}\left(\mathbf{P}_{k,i}\mathbf{Q}_{j,:,i}^{T}-\overline{c}_{i,j,k}^{(1)}\right)^{2}_{Sequentiality}+\lambda_1 \sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{k=0}^{K}\left(\mathbf{S}_{i,:,k}\mathbf{T}_{j,,,k}^{T}-\overline{c}_{i,j,k}^{(2)}\right)^{2}_{Sequentiality}+\lambda\|\theta\|^{2}}_{},
\end{aligned}
\tag{7}
$$

$$
S.t: \quad \mathbf{P}\geq 0,\ \mathbf{Q}\geq 0,\ \mathbf{T}\geq 0 \text{ and } \mathbf{S}\geq 0. \tag{8}
$$

It should be noted that in the completion process, due to the role of the masking tensor M, only observations are used to train the model. However, to simplify the objective function, in formulas (7) and (8), we omit the masking tensor. According to formula (1), we can find that in the original tensor, a known low-voltage station area electricity data element will affect the four-parameter tensors. In addition, the predicted value is not only the combination of the results of the two decomposition modules but also the constraint of local limitation; that is, the predicted value of the low-pressure station area data is affected by the real value related to it and the predicted value of the surrounding time interval at the same time. The matrix extracted from the original tensor has the problem of data sparsity, which can be solved by using nonnegative matrix factorization, and its nonnegativity ensures the interpretability of the learned parameter tensor.

After clarifying the objective function of LPZ, we will next introduce in detail how to estimate the parameters. Compared with the stochastic gradient descent method [18], the alternating least squares method is easier to adjust and highly parallelizable [19]. Therefore, in this article, we refer to the alternating least squares method for parameter estimation. During the training process, the observations will be used to update the model parameters iteratively until the objective function converges. In one iteration, all samples in the original tensor will be traversed once, and at the same time, the four-parameter tensors will be updated with one sample each time. Here, a known element in the tensor is a training sample.

The parameters to be trained are the four-parameter tensors of the model. Here, we separately calculate the objective function for each vector $\{\mathbf{P}_{k,:,i}, \mathbf{Q}_{j,:,i}, \mathbf{S}_{i,:,k}, \mathbf{T}_{j,:,k}\}$. The partial derivative and the result are shown in the following formulas:

$$
\begin{aligned}
\frac{\partial \ell}{\partial P_{k,:,i}} = & 2\beta \sum_{j=1}^{J}\left(\mathbf{P}_{k,:,i}\mathbf{Q}_{j,:,i}^{T}-c_{i,j,k}\right)-Q_{j,:,i} \\
& + 2\lambda_1 \sum_{j=1}^{J}\left(\mathbf{P}_{k,:,i}\mathbf{Q}_{j,:,i}^{T}-\overline{c}_{i,j,k}^{(1)}\right)\mathbf{D}_{j,:,i}^{(1)}+2\lambda\mathbf{P}_{k,:,i},
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
\frac{\partial \ell}{\partial \mathbf{Q}_{j,:,i}} = & 2\beta \sum_{k=1}^{K}\left(\mathbf{P}_{k,:,i}\mathbf{Q}_{j,:,i}^{T}-c_{i,j,k}\right)\left(-\mathbf{P}_{\mathrm{k},:,i}\right) \\
& + 2\lambda_1 \sum_{k=1}^{K}\left(\mathbf{P}_{k,:,i}\mathbf{Q}_{j,:,i}^{T}-\overline{c}_{i,j,k}^{(1)}\right)\mathbf{P}_{k,:,i}+2\lambda\mathbf{Q}_{j,:,i},
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
\frac{\partial \ell}{\partial S_{i,:,k}} = & 2(1-\beta)\sum_{j=1}^{J}\left(\mathbf{S}_{i,:,k}\mathbf{T}_{\mathrm{j},:,k}^{T}-c_{i,j,k}\right)\left(-\mathbf{T}_{j,:,k}\right) \\
& + 2\lambda_1 \sum_{j=1}^{J}\left(\mathbf{S}_{i,:,k}\mathbf{T}_{j,:,k}^{T}-\overline{c}_{i,j,k}^{(2)}\right)\mathbf{D}_{j,:,i}^{(2)}+2\lambda S_{i,:,k},
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
\frac{\partial \ell}{\partial T_{j,:,k}} = & 2(1-\beta)\sum_{i=1}^{I}\left(\mathbf{S}_{i,:,k}\mathbf{T}_{\mathrm{j},:,k}^{T}-c_{i,j,k}\right)\left(-\mathbf{S}_{i,:,k}\right) \\
& + 2\lambda_1 \sum_{i=1}^{I}\left(\mathbf{S}_{i,:,k}\mathbf{T}_{\mathrm{j},:,k}^{T}-c_{i,j,k}^{T}\right)\mathbf{S}_{i,:,k} \\
& + 2\lambda T_{j,:,k}.
\end{aligned}
\tag{12}
$$

Among them, $\mathbf{D}_{j,:,i}^{(1)}$ and $\mathbf{D}_{j,:,i}^{(2)}$ are auxiliary variables. The tensor $\mathbf{D}^{(1)} \in R^{J\times F\times I}$ corresponds to $\mathbf{Q}$. $\mathbf{D}_{j,:,i}^{(1)} \in R^{F}$ is the difference between the expression of time interval j and the mean value of the surrounding time interval. Its formula is shown in (13). Similar to $\mathbf{D}^{(1)}$, the tensor $\mathbf{D}^{(2)}$ is also used to describe the difference in time interval expression, and $\mathbf{D}_{j,:,k}^{(2)} \in \mathbf{R}^{F}$. The difference is that $\mathbf{D}^{(2)} \in \mathbf{R}^{J\times F\times I}$ corresponds to $\mathbf{T}$, and the time interval expression used by $\mathbf{D}_{j,:,k}^{(2)}$ is

generated based on spatial correlation. The formulaic representation is shown in (14).

$$\mathbf{D}_{j,:,i}^{(1)} = \mathbf{Q}_{j,:,i} - \frac{1}{2W}\sum_{\omega=1}^{W}\left(\mathbf{Q}_{j-\omega,:,i} + \mathbf{Q}_{j+\omega,:,i}\right), \tag{13}$$

$$\mathbf{D}_{j,:,k}^{(2)} = \mathbf{T}_{j,:,k} - \frac{1}{2W}\sum_{\omega=1}^{W}\left(\mathbf{T}_{j-\omega,:,k} + \mathbf{T}_{j+\omega,:,k}\right), \tag{14}$$

where W is the window size, and $\omega = 1, 2, \ldots, W$, that is, the power data of a certain time interval will be affected by the time interval before and after it.

The process of parameter update is iterative. Referring to the update process of the alternating least squares matrix decomposition, we set the gradient to 0 to derive the parameter update formula, as shown in the following equations:

$$\mathbf{P}_{k,:,i} = \left[\beta\mathbf{Q}_i^T\mathbf{Q}_i + \lambda_1\mathbf{D}_i^{(1)T}\mathbf{D}_i^{(1)} + \lambda\mathbf{I}\right]^{-1} \cdot \left[\beta\sum_{j=1}^{J} c_{i,j,k}\mathbf{Q}_{j,:,i}\right], \tag{15}$$

$$\mathbf{Q}_{j,:,i} = \left[\beta\mathbf{P}_i^T\mathbf{P}_i + \lambda_1\mathbf{P}_i^T\mathbf{P}_i + \lambda\mathbf{I}\right]^{-1} \cdot \left[\beta\sum_{k=1}^{K} c_{i,j,k}\mathbf{P}_{k_i,i} + \lambda_1\sum_{k=1}^{K}\overline{c}_{i,j,k}^{(1)}\mathbf{P}_{k,:,i}\right], \tag{16}$$

$$S_{i,:,k} = \left[(1-\beta)\mathbf{T}_k^T\mathbf{T}_k + \lambda_1\mathbf{D}_k^{(2)T}\mathbf{D}_k^{(2)} + \lambda\mathbf{I}\right]^{-1} \cdot \left[(1-\beta)\sum_{j=1}^{J} c_{i,j,k}\mathbf{T}_{j,:,k}\right], \tag{17}$$

$$\mathbf{T}_{j,:,k} = \left[(1-\beta)\mathbf{S}_k^T\mathbf{S}_k + \lambda_1\mathbf{S}_k^T\mathbf{S}_k + \lambda\mathbf{I}\right]^{-1} \cdot \left[(1-\beta)\sum_{j=1}^{J} c_{i,j,k}\mathbf{S}_{i,:,k} + \lambda_1\sum_{i=1}^{I}\overline{c}_{i,j,k}^{(2)}\mathbf{S}_{i,:,k}\right]. \tag{18}$$

Specifically, in an iterative update, the tensors **Q**, **S**, and **T** are fixed first, and the row vector of the tensor **P** is updated according to the above formula. After updating of the **P** is completed, **P**, **S**, and **T** are fixed to update **Q** row by row. Next, **S** and **T** are updated separately in this update mode. Obviously, because for an update of a parameter tensor, the update between the row vectors does not affect each other, this process is highly parallelizable, which can greatly speed up the training speed of the model. In one iteration, the four-parameter tensors will be updated separately according to formulas (15)–(18) and continue until the objective function converges. It is worth noting that this update method does not guarantee the nonnegativity of the parameter tensors **P**, **Q**, **S**, and **T**. Because our objective function is continuous, its minimum value should be obtained at the point where the gradient is 0 or the point on the boundary. In this paper, a simple method is used to deal with negative values in the parameter tensor. If there is a value less than 0 in the parameter tensor, set it to 0.

Figure 4 summarizes the training process of the LPZ model. First, the original low-voltage station area electricity data tensor is formed and used as the input of the model. Then, the model is trained and the four-parameter tensors are iteratively updated until the objective function converges (lines 2–8). Finally, the results of the decomposition modules are averaged in a weighted manner and used as the predicted values of the missing values (lines 9–13).

4. Experiment and Analysis

For the validity of our model, we conducted a large number of experiments on a dataset of a certain station in a distribution network of a certain city in China. The LPZ is compared with three current data complementation methods, and the experimental results show that LPZ can obtain better prediction results than the current complementation methods. In this section, we first introduce the dataset and experimental settings; second, we use different parameters to evaluate the LPZ model; and finally, we conduct comparative experiments and analyse the experimental results.

4.1. Dataset and Experimental Settings. This section first introduces the dataset used in the experiment and then explains the experimental parameter settings and evaluation indicators.

4.1.1. Dataset. For the dataset of this experiment, we will use the user current data of a certain station in the distribution network of a certain city in China. The station structure is shown in Figure 5. Here, VLV22 represents the cable model, and 4×70 represents 4-core 70 mm^2. In the actual data collected automatically, the current data of a certain month are randomly selected as the data test set to construct the current tensor. The constructed current tensor contains a total of 142650 ($317\times180\times25$) elements, of which the number of nonzero elements is 1209011. We randomly selected 80% of the nonzero elements as the training set and the remaining 20% as the test set to prove the effectiveness of our proposed model.

4.1.2. Parameter Setting. In this section, the parameter settings of the LPZ model are mainly discussed, and these default settings are obtained through parameter tuning. In the experiment, we set the default value of the number of hidden factors **F** to 15; that is, we use a 15-dimensional vector to represent users, time intervals, and days. The weight parameter β is used to control the combining process of the partial results of the two subdecomposition modules, and its default setting is 0.4. The default window size W is 4; that is, the current data of a time interval are affected by 4-time intervals before and after it. In the local limit, λ_1 is used to control the influence of sequentiality, and the default value is 0.1. At the same time, for the regularization term coefficient λ, the default value is also set to 0.1.

Algorithm 1: CIM learning algorithm

Input: The original low-pressure station data tensor with missing values $\boldsymbol{C}$, mask tensor $\boldsymbol{M}$, number of hidden factors $\boldsymbol{F}$ and the parameter tensors $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{S}$ *and* $\boldsymbol{T}$;

Output: Fully estimated tensor $\widetilde{\boldsymbol{C}}$;

Initialize $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{S}$ *and* $\boldsymbol{T}$; // Training model

repeat

for $c_{i,i,k} \in \boldsymbol{C}$ **do**

if $m_{i,j,k} = 0$ **then**

Update the parameter tensor according to formula (15)-(18);

end if

end for

until the stop condition // prediction is met

for $m_{i,j,k} \in M$ **do**

if $m_{i,j,k} = 1$ **then**

Calculate missing values according to formula (4)

end if

end for

return fully estimated tensor $\widetilde{\boldsymbol{C}}$;

Figure 4: Training algorithm of the LPZ model.

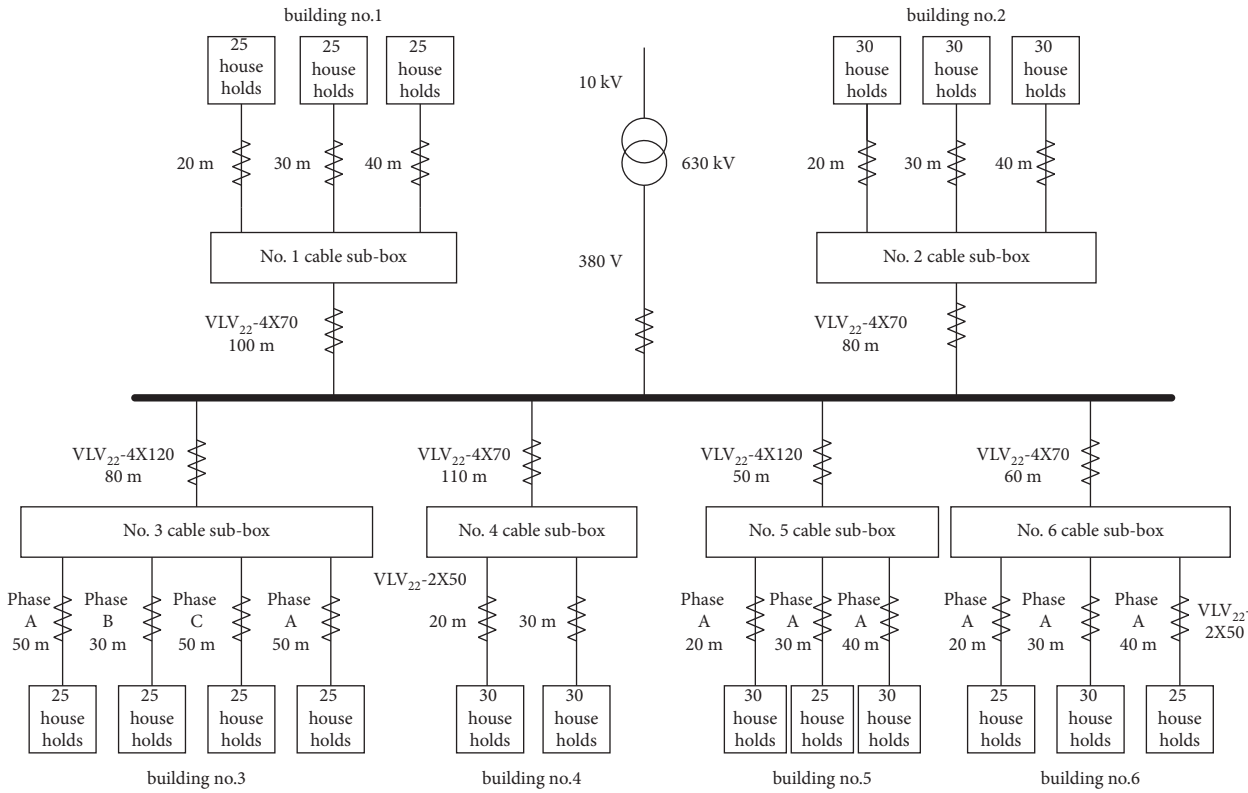


Figure 5: Structure diagram of the low-voltage station area.

4.1.3. Evaluation Index. In this article, we use root mean square error (RMSE) and mean absolute error (MAE) as evaluation indicators to evaluate the experimental results [20, 21]. Given a sparse low-voltage station area data tensor, N represents the total amount of missing data in this tensor, that is, the size of the test set in this experiment. The formula of the evaluation index is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} M_{i,j,k}\left(c_{i,j,k} - \widetilde{c}_{i,j,k}\right)^2}, \tag{19}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} m_{i,j,k}c_{i,j,k} - \widetilde{c}_{i,j,k}, \tag{20}$$

where $c_{i,j,k}$ represents the true value and $\widetilde{c}_{i,j,k}$ represents the predicted value. To make the evaluation more convenient, the mask $m_{i,j,k} \in \{0, 1\}$.

4.2. Experimental Results and Analysis. We consider that in the case of random missing electricity data in low-voltage area and long-term missing data due to faults, this paper is verified by analysing the performance of the proposed method under different time granularities and different training set sizes and the results of long-term missing data completion. Based on the effectiveness of the proposed method and comparing the method in this article with three missing current data complementation methods, it is concluded that the method proposed in this article is better than other methods.

4.2.1. Random Missing Data Completion. To verify the performance of our model, the time interval size was set to 5 minutes, 10 minutes, and 15 minutes, and the experimental results at different time granularities were obtained, as shown in Table 1.

It can be seen from the table that as the time interval increases, the evaluation indicators (RMSE and MAE) of the proposed model show a downward trend, and the complement effect continues to improve. This is because as the time interval increases when predicting missing values, more users' power consumption conditions are taken into account, and the prediction effect will be better.

In addition, we also changed the size of the training set for experiments. Taking $\Delta t = 15$ minutes as an example, we randomly selected 30%, 50%, 70%, and 90% of the original training set as the new training set for experimentation. The experimental results are shown in Table 2. It can be seen from the table that the complete performance of the proposed model increases with the increase of the training set. This is because the larger the training set is, the more the available samples can be used to mine user current data information, thereby obtaining more information and leading to accurate completion results. Therefore, the method in this paper can effectively complete data completion in practice.

4.2.2. Completion of Missing Data throughout the Day. A serious failure may occur during the collection or transmission of the power data by the smart meter, which cannot be recovered in a short time, and the user power data may be lost for a whole day [22]. This paper verifies the effectiveness of the proposed method by randomly discarding current data for several days.

Figure 6(a) shows the actual data and completion results when a user's data are missing for 7 days throughout the day. It can be seen from the figure that the two curves basically overlap, indicating that the completion effect is better. Figure 6(b) is the area chart of the difference between the complement value and the actual value when the entire day is missing for 7 days. It can be seen from the figure that the missing data can still be prepared for the missing data in the entire day.

Table 3 shows the performance data of this method under different missing days. It can be seen from the table that the changes in RMSE and MAE increase with the increase in missing days. When the number of missing days is less than 12 days, both RMSE and MAE are relatively small, indicating that the method proposed in this paper can compensate for the missing days. The entirety is still valid, and the fewer the missing days there are, the better the completion effect. In practice, the missing data will generally be repaired within a week, and the lack of more days rarely happens. Therefore, the method in this article is also more applicable in practice.

4.2.3. Experimental Comparison. The method in this article is compared with three missing current data completion methods (cubic spline interpolation [23], Kalman filtering [24], and tensor completion [25]). Figure 7 shows that the 31-day use of the dataset in the station area is the original complete data. The average absolute error and root mean square error trend are set at 40%–100%. It can be seen from the figure that the completion errors of all methods decrease with the increase of the dataset used. In addition, the accuracy of the station data completion of different methods is also different. The completion error curves (RMSE and MAE) of the three methods of Kalman filtering, cubic interpolation, and tensor completion are all above the joint matrix decomposition completion error curve. That is, when the joint matrix factorization and completion method has the same dataset size, its complete accuracy is higher than that of the other three methods. It is worth noting that the two types of errors are complemented by the joint matrix factorization method; when the dataset size is 50% of the original complete dataset, the error is only approximately equal to 10% of the error of other methods, indicating that the joint matrix factorization method is complementary. Compared with other methods, the whole method is more suitable for the completion of high-deficiency cases.

Figure 8 shows the variation trend of the root mean square error and the mean absolute error of the 31-day data of randomly missing all days in the station area from 1 to 12 days. It can be seen from the figure that the errors of all-day missing data completion of all the completion methods

Table 1: Experimental completion error at the same time granularity.

Time granularity Δt (min)	RMSE/A	MAE/A
5	0.3615	0.2238
10	0.3447	0.2065
15	0.3365	0.1975

Table 2: Experimental completion errors under different training set sizes.

Training set size (%)	RMSE/A	MAE/A
30	0.6615	0.2725
50	0.4647	0.2038
70	0.3065	0.1315
90	0.1005	0.0538

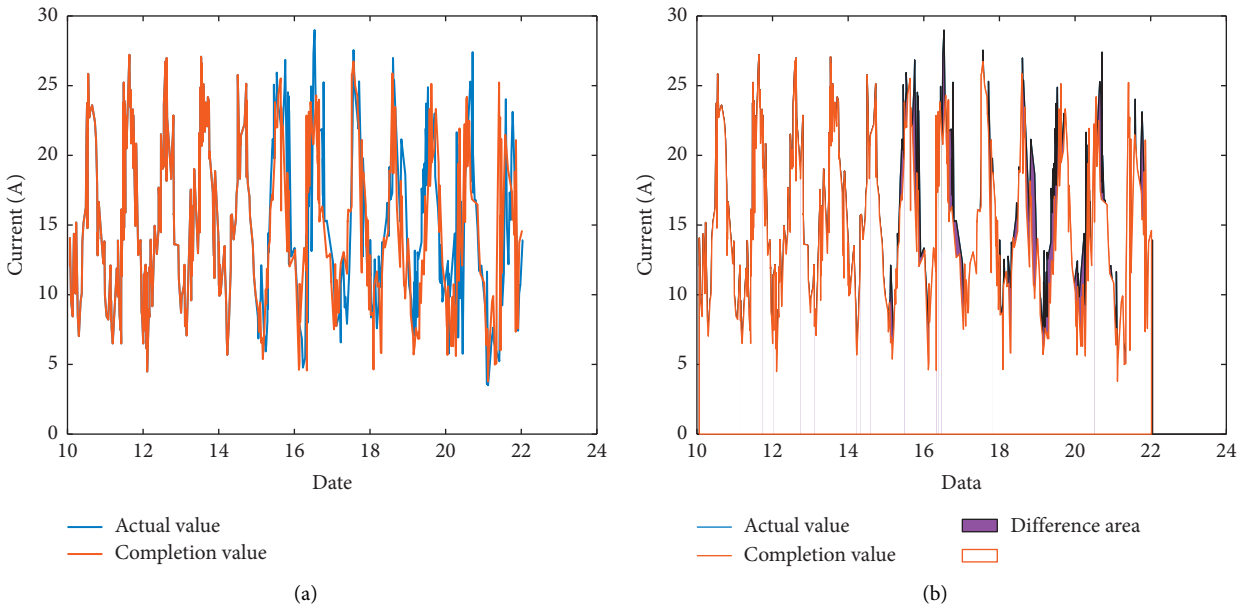


Figure 6: Complete map of the experiment with 7 days missing all day. (a) Completion curve when seven days are missing in the whole day. (b) The area of the difference between the true value and the complement value when the entire day is missing for seven days.

Table 3: Experimental completion error under different missing days.

Missing days/d	RMSE/A	MAE/A
1	0.8615	0.1725
4	1.2047	0.6438
8	1.7765	1.2315
12	2.3605	2.2238
16	4.8725	2.6318
20	6.3215	4.3159
24	7.2638	5.1245

increase with the increase of the number of missing days, and the two error curves of the joint matrix decomposition completion are both below those of cubic interpolation, Kalman filtering, and tensor compensation. The full bottom, that is, the precision, of joint matrix factorization is higher than that of other methods.

Through the comparison, it can be seen that the joint matrix factorization completion method has better

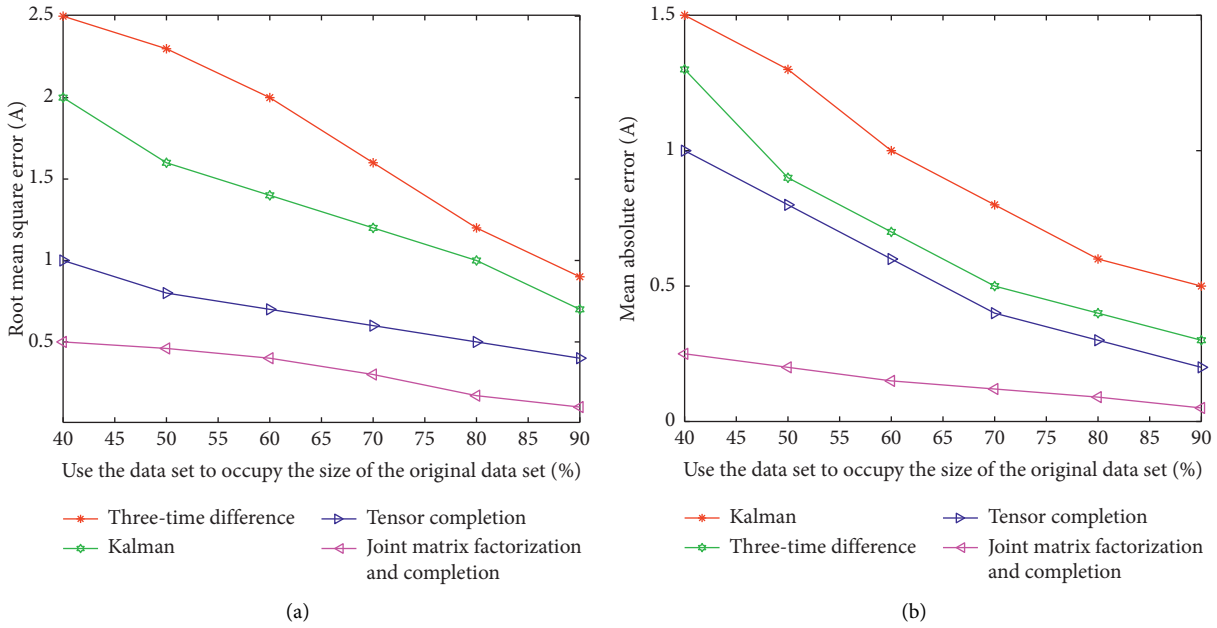


Figure 7: Error comparison results of different dataset sizes. (a) Root mean square error. (b) Mean absolute error.

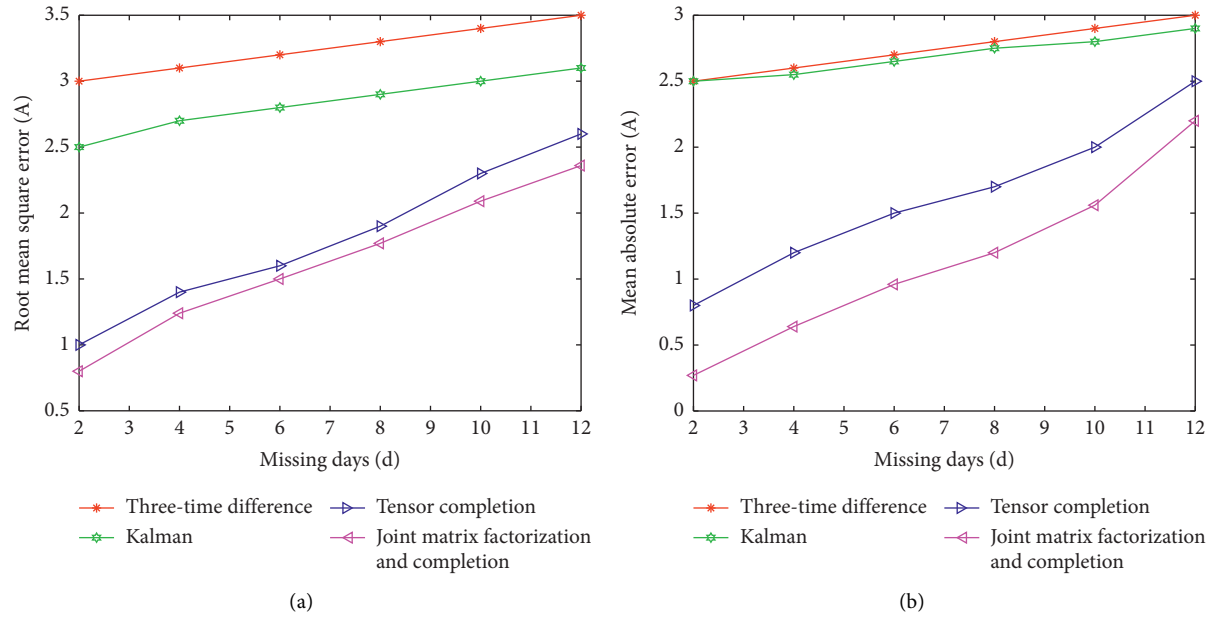


Figure 8: Error comparison results under different missing days. (a) Root mean square error. (b) Mean absolute error.

completion effects than the cubic interpolation, Kalman filter, and tensor completion methods in the case of random missing data or all-day missing data.

4.3. Model Parameter Tuning. In this section, the influence of important parameters on the performance of the model is evaluated, and the experimental results are analysed under different parameter values. In our model, the main focus is on two decomposition modules and local restrictions. Therefore, the main parameters of our research include the number of hidden factors (F), periodic weight parameters (β), and window size ($\mathbf{W}$) and local restriction weight (λ_1).

Figures 9(a) and 9(b) respectively show the changes of RMSE and MAE with the number of latent factors F. RMSE and MAE gradually decrease with the increase of F. This is because the vector in the high-dimensional space can better reflect the influence relationship between different modes. However, when F is too large, the performance of the model begins to degrade. This is because the number of redundant parameters that the model learns is too large due to limited observations, causing the model to overfit. In addition, it can

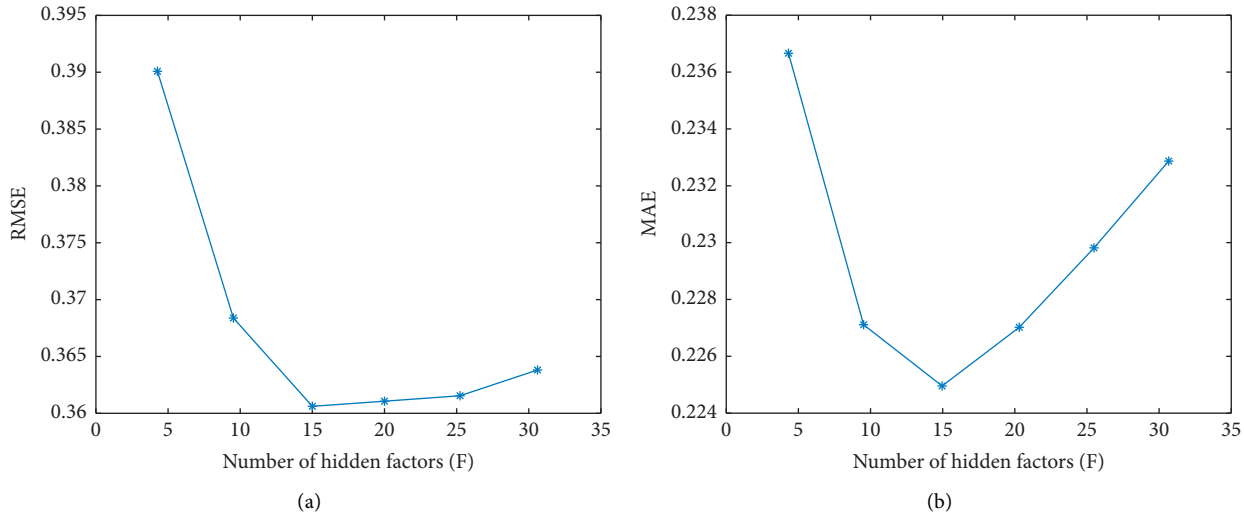


Figure 9: The influence of the number of hidden factors F on the experimental results. (a) RMSE. (b) MAE.

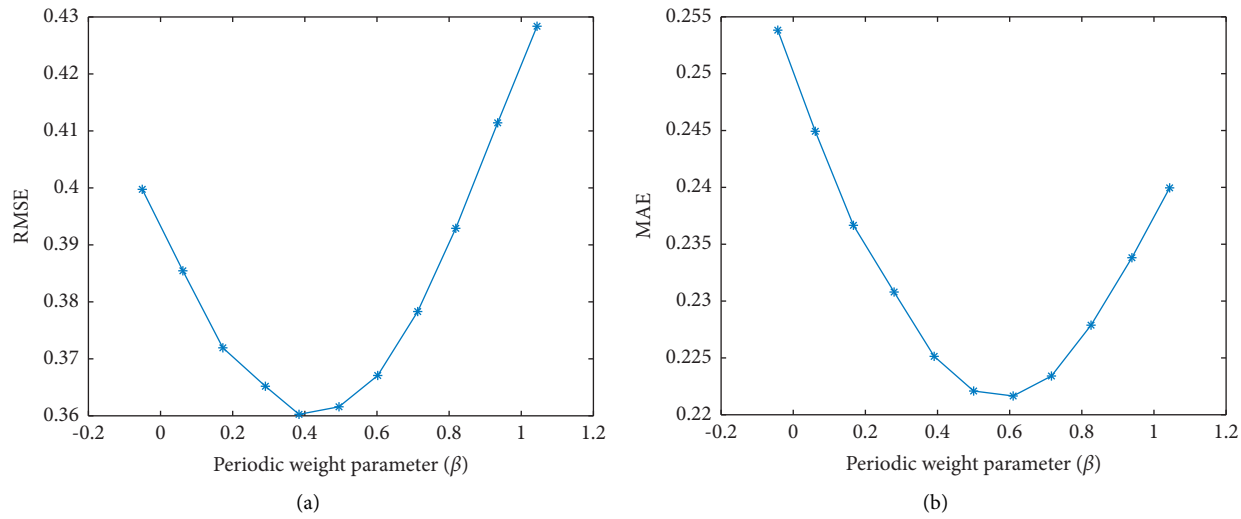


Figure 10: The influence of the weight parameter β on the experimental results. (a) RMSE. (b) MAE.

be seen that the optimal value of RMSE is obtained when F = 15.

In order to explore the influence of the periodic weight parameter β on the model, the value of β is gradually changed from 0 to 1, while the step size is 0.1. The experimental results are shown in Figure 10. A larger value of β contributes to stronger influence of the relationship between the day and the time interval in the model at this time and better performance of the model. Therefore, as β increases, RMSE and MAE gradually become smaller. However, when the value of β exceeds 0.4, the performance of the model begins to gradually decline. This is because the relationship between the day and the time interval is overemphasized in the forecasting process, while the relationship between the user and the time interval is ignored. Here, it is worth noting that $\beta = 0$ means that only periodicity plays a role; when $\beta = 1$, only spatial correlation is used. From Figure 10, we can find that the performance of LPZ can be significantly improved by considering both periodicity and spatial correlation.

In addition, the value of the window size **W** is changed to observe its influence on the model performance. The experimental results are shown in Figure 11. In general, the performance of the model will be improved as **W** increases, because increasing **W** means that more samples can be used to learn temporality. However, RMSE and MAE begin to gradually decrease, when **W** is greater than 4. This phenomenon is caused by the characteristics of timing, that is, the correlation between low-voltage station data at different time intervals will be weakened as the distance increases. Therefore, the time span should be selected appropriately to predict the low-pressure station data at a certain time interval. Too large or too small time granularity of the division will affect the accuracy of the model's prediction results.

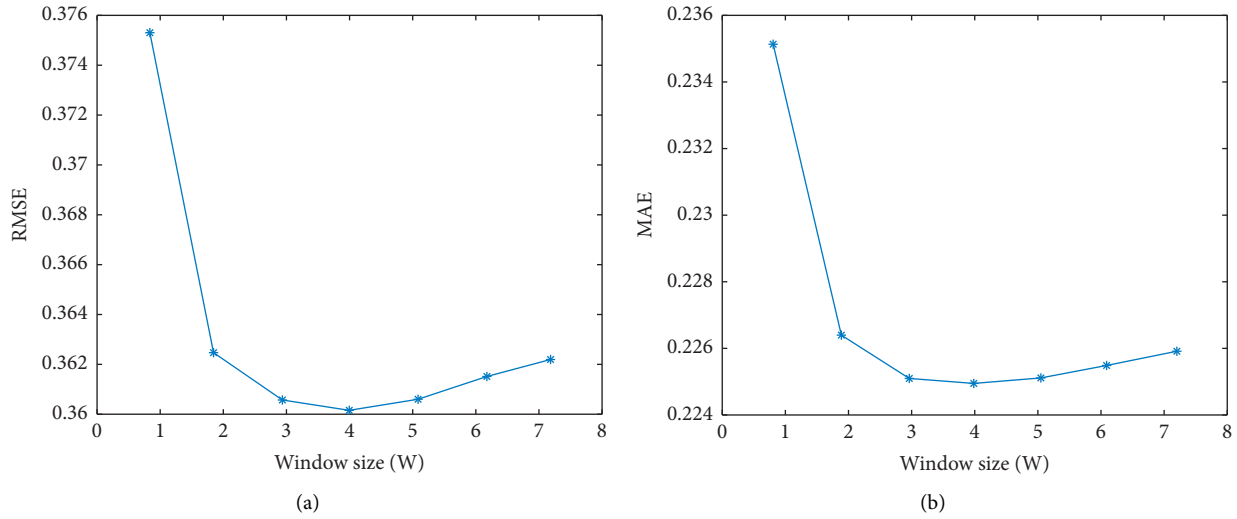


Figure 11: The influence of window size **W** on experimental results. (a) RMSE. (b) MAE.

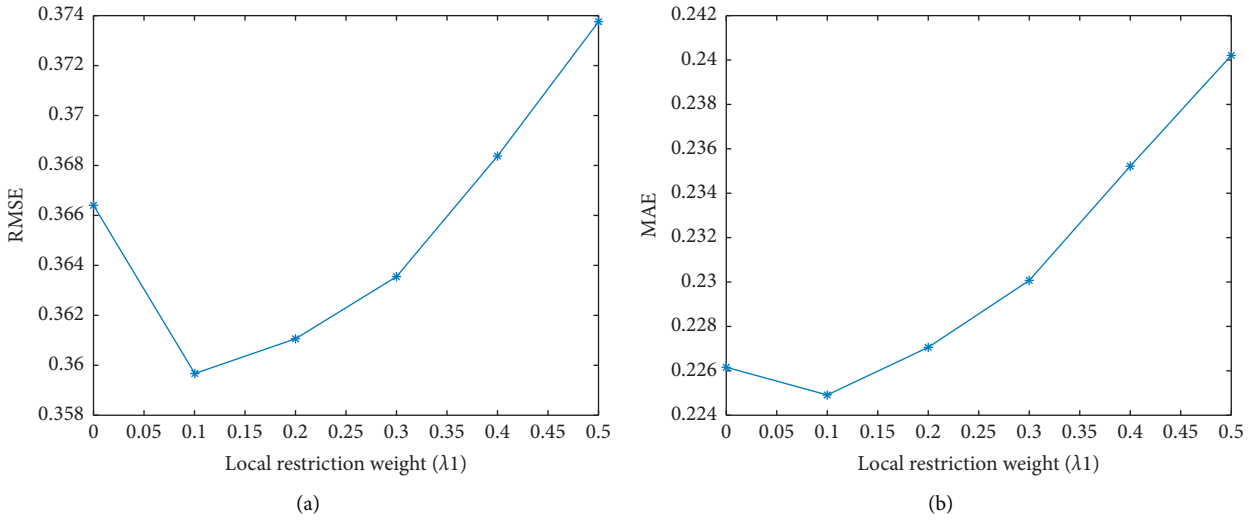


Figure 12: The influence of the local limit parameter λ_1 on the experimental results. (a) RMSE. (b) MAE.

Finally, the performance of the model is analysed under different local restriction weights. As shown in Figure 12, the model performs obviously worse at $\lambda_1 = 0$ than $\lambda_1 = 0.1$, because no local restriction is added to the decomposition module at $\lambda_1 = 0$. This shows that introduction of the time dependency of the relative travel time to the model can improve the performance of the model. When λ_1 is greater than 0.1, the model evaluation index increases with the continuous increase of λ_1. This indicates that we cannot pay too much attention to the timing. It can be seen from Figures 11 and 12 that simultaneous consideration of the periodicity, spatial correlation, and timing of low-voltage station data can improve the performance of LPZ in the task of missing value completion.

5. Conclusions

This paper proposes a model that jointly considers the periodicity, time series, and spatial correlation of the electricity data in the low-voltage station area for the problem of the lack of supplementary power consumption data in the low-voltage station area. The model proposed in this paper can reasonably estimate unknown data based on the observation value of limited low-voltage station area electricity data. Aiming at the characteristics of the electricity data in the low-voltage station area, we fold and stack the electricity data sequence into a three-dimensional tensor form and use it as the input of the joint matrix decomposition module. In the decomposition module, we not only model the

periodicity and spatial correlation but also add local restrictions to make it subject to timing constraints. To verify the effectiveness of the model, we conducted many experiments on a real dataset and compared LPZ with traditional completion methods (cubic interpolation, Kalman filtering, and tensor completion methods). Experiments show that LPZ can not only complement the missing power data but also that the effect of the complement is better than that of traditional complement methods.

To sum up, the completion method based on joint matrix decomposition proposed in this paper is not only suitable to solve the problem of data missing in low-voltage stations but also other data missing problems. With a wide range of application prospects, this method has the advantage of being less affected by the length of the missing data and the location of the missing data. To a certain extent, improving the accuracy of data completion can contribute to higher accuracy of data mining and analysis. Data completion is essential for data mining and analysis and is also of practical value for load forecasting in the power grid and even for power generation forecasting such as photovoltaic power generation and wind power generation.

Data Availability

The experimental data are obtained by Python simulation, and the experimental result diagram is obtained by MATLAB.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Y. Wadhawan, A. AlMajali, and C. Neuman, "A comprehensive analysis of smart grid systems against cyber-physical attacks," *Electronics*, vol. 7, no. 10, p. 249, 2018.

[2] Y. Zhang, Y. Ting, and M. Guangyu, "Review and prospect of ubiquitous power Internet of things in smart distribution system," *Electric Power Construction*, vol. 40, no. 6, pp. 1–12, 2019.

[3] G. Zhang and J. Guo, "A novel method for hourly electricity demand forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1351–1363, 2020.

[4] S. Oprea, "Data framework for electricity price setting in competitive environment," *Journal of Physics: Conference Series*, vol. 1297, 2019.

[5] M. Zhou, Y. Li, M. J. Tahir, X. Geng, Y. Wang, and W. He, "Integrated statistical test of signal distributions and access point contributions for Wi-Fi indoor localization," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 5057–5070, 2021.

[6] J. Wang, X. Wang, C. Ma, and L. Kou, "A survey on the development status and application prospects of knowledge graph in smart grids," *IET Generation, Transmission & Distribution*, vol. 15, no. 3, pp. 383–407, 2021.

[7] H. Yang, Y. De Ng, and Z. Liu, "Study on electric load forecasting with historical bad data," *Dianli Xitong Baohu yu Kongzhi/Power System Protection and Control*, vol. 45, no. 15, pp. 62–68, 2017.

[8] M. Gurusamy and P. Vijayakumar, "An efficient cloud data center allocation to the source of requests," *Journal of Organizational and End User Computing*, vol. 32, no. 3, pp. 23–36, 2020.

[9] Q. Liu, X. Li, and H. Cao, "Two-dimensional localization: low-rank matrix completion with random sampling in massive MIMO system," *IEEE Systems Journal*, vol. 15, 2020.

[10] P. Gao, M. Wang, and S. G. Ghiocel, "Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1006–1013, 2015.

[11] M. Yang, Y. Sun, and G. Mu, "Data completing of missing wind power data based on adaptive neuro-fuzzy inference system," *Automation of Electric Power Systems*, vol. 38, no. 19, pp. 16–21, 2014.

[12] G. Tutz and S. Ramzan, "Improved methods for the imputation of missing data by nearest neighbor methods," *Computational Statistics and Data Analysis*, vol. 90, 2015.

[13] H. Gu, T. Wang, Y. Zhu, C. Wang, D. Yang, and L. Huang, "A completion method for missing concrete dam deformation monitoring data pieces," *Journal of Applied Sciences*, vol. 11, no. 1, 2021.

[14] M. Zhou, Y. Long, W. Zhang et al., "Adaptive genetic algorithm-aided neural network with channel state information tensor decomposition for indoor localization," *IEEE Transactions on Evolutionary Computation*, vol. 25, no. 5, pp. 913–927, 2021.

[15] L. Li, J. Zhang, and Y. Wang, "Missing value imputation for traffic-related time series data based on a multi-view learning method," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 2933–2943, 2018.

[16] V. Salerno and G. Rabbeni, "An extreme learning machine approach to effective energy disaggregation," *Electronics*, vol. 7, no. 10, p. 235, 2018.

[17] F. Meng, Q. Ji, H. Zheng, H. Wang, and D. Chu, "Modeling and solution algorithm for optimization integration of express terminal nodes with a joint distribution mode," *Journal of Organizational and End User Computing*, vol. 33, no. 4, pp. 142–166, 2021.

[18] V. Gandikota, D. Kane, and R. K. Maity, "Vqsgd: vector quantized stochastic gradient descent," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 2197–2205, San Diego, CA, USA, April 2021.

[19] H. Duan, X. Xiao, J. Long, and Y. Liu, "Tensor alternating least squares grey model and its application to short-term traffic flows," *Applied Soft Computing*, vol. 89, p. 106145, 2020.

[20] M. Ćalasan, S. H. E. A. Aleem, and A. F. Zobaa, "On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: a novel exact analytical solution based on Lambert W function," *Energy Conversion and Management*, vol. 210, p. 112716, 2020.

[21] J. Qi, J. Du, S. M. Siniscalchi, X. Ma, and C.-H. Lee, "On mean absolute error for deep neural network based vector-to-vector regression," *IEEE Signal Processing Letters*, vol. 27, pp. 1485–1489, 2020.

[22] A. H. Afsharinejad, C. Ji, and R. Wilcox, "Heterogeneous recovery from large scale power failures," 2020, https://arxiv.org/abs/2012.15420.

[23] S. Huang, J. Tang, and J. Dai, "1DCNN fault diagnosis based on cubic spline interpolation pooling," *Shock and Vibration*, vol. 202013 pages, 2020.

[24] D. Liu, Z. Wang, Y. Liu, and F. E. Alsaadi, "Extended Kalman filtering subject to random transmission delays: dealing with

packet disorders," *Information Fusion*, vol. 60, pp. 80–86, 2020.

[25] C. Cai, G. Li, and H. V. Poor, "Nonconvex low-rank tensor completion from noisy data," 2019, https://arxiv.org/abs/1911.04436.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 4500957, 11 pages
https://doi.org/10.1155/2021/4500957

Research Article

Hyperspectral Image Denoising Based on Nonconvex Low-Rank Tensor Approximation and l_p Norm Regularization

Li Bo,[1] **Luo Xuegang**,[2] **and Lv Junrui**[2]

[1]*School of Information and Engineering, Sichuan Tourism University, Chengdu, Sichuan 610100, China*
[2]*School of Mathematics and Computer Science, Panzhihua University, Panzhihua 617000, China*

Correspondence should be addressed to Li Bo; sctulibo2021@126.com

Received 24 September 2021; Accepted 16 October 2021; Published 5 November 2021

Academic Editor: Xianyong Li

A new nonconvex smooth rank approximation model is proposed to deal with HSI mixed noise in this paper. The low-rank matrix with Laplace function regularization is used to approximate the nuclear norm, and its performance is superior to the nuclear norm regularization. A new phase congruency l_p norm model is proposed to constrain the spatial structure information of hyperspectral images, to solve the phenomenon of "artificial artifact" in the process of hyperspectral image denoising. This model not only makes use of the low-rank characteristic of the hyperspectral image accurately, but also combines the structural information of all bands and the local information of the neighborhood, and then based on the Alternating Direction Method of Multipliers (ADMM), an optimization method for solving the model is proposed. The results of simulation and real data experiments show that the proposed method is more effective than the competcing state-of-the-art denoising methods.

1. Introduction

Due to the influence of many factors in the process and transmission of hyperspectral images, the acquired hyperspectral images often contain some complex mixed noises, including gauss noise, salt-and-pepper noise, and dead-line noise. It is very difficult to analyze and apply the hyperspectral image to high-level applications. In the early days, a great number of methods were proposed to remove noise, such as Fourier Transform, Wavelet Transform, nonlocal means (NLM) filter, block-matching, and 3D filtering (BM3D). However, most of the above methods require some prior knowledge of noise and can only deal with one or two types of noise. In addition, HSI data in the real-world are often mixed into the real data by various noise combinations. To solve this problem, some multidimensional methods are proposed to deal with both spectral and spatial information. In [1], a multidimensional wiener filtering (MWF) algorithm was proposed, which represents the histogram of an image as a three-dimensional tensor and uses tensor analysis to remove the noise. In addition, spectral and spatial information [2] 3D wavelets have obtained good performance.

In recent years, low-rank Matrix Recovery (LRMR) [3] has been used in HIS denoising. Different from the traditional method, LRMR can process different types of noise without any prior information of noise; many methods of HSI image denoising [4–7] based on low rank and spectral correlation have been proposed. Due to the difficulty in solving the low-rank constraint model directly, the kernel norm is used to approximate the low-rank model, and the good results of HSI mixed noise removal are obtained. The representative methods include weighted low-rank model (WLRM) [8], rank minimization (RM) [9], structure tensor total variance weighted nuclear norm minimization (STTV-WNNM) [10], weighted nuclear norm minimization (WNNM) [11], and nonlocal low-rank approximation (NLRA).

These methods mainly have taken advantage of the low-rank feature of low-rank description image with nuclear norm approximation and preserving edge structure of image with total variation regularization. However, the low-rank model based on nuclear norm approximation is not a good approximation rank function but lacks image edge sparsity and structure smoothness regularization. If NSS prior

cannot be well used to represent the structure of the image in the low-rank model, the important structure will inevitably be lost, and an "artificial artifact" will appear.

Therefore, it is necessary to approach the rank function better. To overcome the above limitation, the method based on nonconvex smooth rank approximation (SRA) is studied in this paper. The key idea is to use nonconvex smooth functions to approximate rank functions directly and to provide a more rigorous approximation than traditional methods.

In addition, the method based on the fractional band total variation regularization has the following defects: (1) The total variation regularization is essentially a first-order partial differential equation, which is a kind of ill-posed first-order inversion problem; if the original image is disturbed, it is possible to have a large oscillation on the partial derivative. (2) The total variation model is based on the assumption that the image is piecewise linear continuous, and the sharp edges of the image can be effectively preserved. However, only the gradients of adjacent pixels are used in the total variation model discretization; it cannot describe the true edge and structure information of the image accurately and effectively, which may lead to the phenomenon of "artificial artifact." Total generalized variation (TGV) [12], high-order total variation (HOTV) [13], and Schatten p-norm constraint models [14] are used to deal with the "artificial artifacts" caused by the total variation model; it is more accurate to describe image edge information in a natural image, but this kind of method is not suitable for hyperspectral image processing.

Aiming at the shortcomings of the existing subband TV regularization and low-rank tensor denoising methods, which cannot effectively utilize the local neighborhood information, it is easy to cause the "artificial artifact" phenomenon, especially in the curved edge. The gradient cannot accurately describe the true structure of the image. Therefore, in this paper, a new phase congruency l_{p} norm constraint is proposed to constrain the spatial structure information of hyperspectral images, to solve the phenomenon of "artificial artifact" in the process of hyperspectral image denoising.

2. State of the Art

2.1. Phase Congruency. Phase congruency (PC) [12] is a phase-based frequency-domain feature detector proposed by Morrone and Owens, which can detect a wide range of visual features and is invariant to local smooth illumination variations. Unlike the gradient in the spatial domain, the PC uses the frequency domain to get the spectrum information at the maximum overlap of the phases, for the high-order edge and corner, line, step edge and roof Ridge, and other visual sensitive important image feature capture more complete. In [12], the phase consistency of signal x is defined by Fourier series expansion shown as

$$PC(x) = \max_{\phi'(x)\in[0,2\pi]} \frac{\sum_n A_n \cos\left(\phi_n(x) - \phi'(x)\right)}{\sum_n A_n}, \quad (1)$$

where A_n is the Fourier series of the n term, and $\phi_n(x)$ is the local phase information of the Fourier series for signal x, and the PC value is the maximum value of equation (1) for the parameter $\phi'(x)$, which is the weighted average of all the amplitudes corresponding to the Fourier terms. Because the PC function defined by equation (1) is greatly influenced by noise, it is easy to produce edge offset, which leads to the loss of part of image structure and contour features.

A phase congruency (MPC) [13] was presented for extracting three orthogonal characteristic components of an image, namely, amplitude, direction, and phase, from a single-pass signal theory and applied to image quality evaluation [14] and image denoising applications, and the satisfactory results of time efficiency and detection effect were obtained.

The phase congruency based on single-pass signal theory is defined as follows:

$$\mathrm{MPC}(x) = W(x)\left\lfloor 1 - \xi \times a\ \cos\left(\frac{E'(x)}{A'(x)}\right)\right\rfloor \frac{\lfloor E'(x) - T\rfloor}{A'(x) + \varepsilon}, \quad (2)$$

where $\mathrm{Weight}(x)$ is the weight function, ξ is an approximate gain factor for sharpening the edge response, and the value range [1, 2], $T = \sum_\theta T_\theta$ is the compensation noise effect. $W(X)$ is defined by

$$W(x) = \frac{1}{(1 + \exp(\gamma(s - c(x))))}, \quad (3)$$

where γ is the gain factor, S is the cut-off value of the filter response expansion, and $c(x)$ is the fractional scatter measure, whose value is divided by the response amplitude and the highest response value $c(x) = A'(x)/(N * (A_{\max}(x) + \varepsilon))$, $\in (0, 1)$, where N is the scale quantity.

Phase consistency l_{p} norm (PCSP) is defined as the Schatten p norm of the PC value on the definition of the singleton phase consistency function in equation (2).

$$\mathrm{PCSP}(Y)_p = \sum_{i=0,j=0}^{MN} \left\|\mathrm{PC}\left(y_{i,j}\right)\right\|_p, \quad (4)$$

where M and N are the height and width of the image, respectively, $P \geq 1$. Equation (4) is difficult to optimize the Schatten p norm due to the existence of convolution kernel operation. To solve this problem, equation (4) needs to be transformed into an equivalent form that is easy to be solved.

Given a Matrix $Y \in \mathbb{R}^{M\times N}$, then the singular value of the Matrix can be decomposed into $Y = U\sum V^T$, where U and V are the left and right singular value eigenvectors of Y, respectively, and the diagonal element σ of the Matrix Y is the singular eigenvalue. The Schatten P norm of a Matrix is defined by

$$\|Y\|_{Sp} = \left(\sum_{i=0}^{\min\{M,N\}} \sigma^p\right)^{1/p}, \quad (5)$$

where σ_{i} is the ith singular eigenvalue of the Matrix Y, corresponding to the (i, i) element value. From equation (5), the PCTV equivalent of equation (4) can be defined by

$$\|\mathrm{PC}(Y)\|_{1,p} = \sum_{i=1}^{\min\{M,N\}} \left\|\sigma_{\mathrm{MPC}}\right\|_{Sp}. \tag{6}$$

2.2. Nonconvex Low-Rank Approximation Functions. Because the nuclear norm cannot approximate the rank function well, it is unreasonable to replace the rank function with the nuclear norm directly. In the field of image restoration, the method of approximating rank minimization by nonconvex substitution has received extensive attention. The common nonconvex functions and corresponding hypergradients are shown in Table 1.

To provide more rigorous approximations than the nuclear norm, we have conducted numerical experiments on one-dimensional data. We selected five experiments of rank approximation of regularization terms with the best performance, as shown in Figure 1. As can be seen from Figure 2, nonconvex optimization is often superior to convex optimization.

As can be seen from Figure 2, the nuclear norm deviates from the true rank, so all singular values are treated equally. The weighted nuclear norm and the Garman norm are the neutralization between the rank minimization and the kernel norm, which can increase the penalty to the small value and decrease the penalty to the large value. The Logdet norm is poor at small singular values, especially those close to 0. The exponential function is used to deal with different singular values, so that Laplace modules and real rank have obvious consistency. To maintain accuracy and speed, the Laplace norm is the best choice for approximating the actual rank of the NSS matrix composed of a similar patch.

In this section, the Laplace function is chosen to approximate the rank function. Compared with other tensor nuclear norms, the Laplace function proxy norm is a better method to measure the rank of the tensor. We introduce the proposed proxy into a low-rank tensor separation model and solve the model by using the ADMM algorithm, which can effectively complete the missing elements in the tensor, to achieve the goal of hyperspectral image denoising. A large number of experiments show that this method is better than the existing methods.

3. HSI Denoising Based on Nonconvex Low-Rank Tensor Approximation

3.1. The Motivation of the Model. The method based on tensor nuclear norm minimization and TV regularization has achieved some noise reduction results in hyperspectral denoising applications, but due to the limitation of TV total variation and the fact that the kernel norm cannot accurately describe the feature of low tensor rank, as a result, its noise removal performance is limited, and it is difficult to meet the demand of hyperspectral image denoising ability for practical applications. Based on the minimization of nonconvex low-rank approximation, the weight of singular value can be adjusted according to the main features of the image, and the global low-rank characteristic of the HSI image can be used well. In addition, phase congruency can describe and capture image visual features more comprehensively than TV constraints based on gradient information. Therefore, in this subsection, nuclear norms are replaced by tensor-based nonconvex proxy functions. Because of the limitation of low-rank feature detection, phase consistency is used to preserve image edge structure. Therefore, a new HSI denoising method based on nonconvex low-rank tensor approximation and phase consistency constraints is proposed in this section.

3.2. Model Description

3.2.1. Constraint Term of Nonconvex Function and Its Properties. Before elaborating on the denoising model, we first review the definition of the normalized Laplace function, which is shown as

$$\|\mathbf{Y}\|_{\mu} = \sum_{k=1}^{B} \sum_{i=1}^{\min(W,H)} g_{ki}\left(Y^{(k)}\right), \tag{7}$$

and $g(\sigma_i(Y^{(k)})) = \mathrm{sgn}(\sigma_i(Y^{(k)}))(1 - e^{\sigma_i(Y^{(k)})/\sigma_1(Y^{(k)})/\mu})$.

Laplace functions have the following useful properties.

Proposition 1. $\lim_{\mu\longrightarrow 0}\|\mathbf{Y}\|_{\mu} = \sum_{k=1}^{B} rank(Y^{(k)})$

Proof. for $g(\sigma_i(Y^{(k)})) = \mathrm{sgn}(\sigma_i(Y^{(k)}))(1 - e^{\sigma_i(Y^{(k)})/\sigma_1(Y^{(k)})/\mu})$

And $\lim_{\varepsilon\longrightarrow 0} g(\sigma_1(\overline{X}^{(K)})) = \begin{cases} 0, & \text{if } \sigma_1(\overline{X}^{(K)}) = 0, \\ 1, & \text{if } \sigma_1(\overline{X}^{(K)}) > 0. \end{cases}$

So $\lim_{\mu\longrightarrow 0}\sum_{i=1}^{\min(W,H)} g(\sigma_1(\overline{X}^{(k)})) = \mathrm{rank}(\overline{X}^{(K)})$ come to the conclusion. □

Proposition 2. $\|UXV\| = \|X\|_{\mu}$,
For any tensor, $U \in \mathbb{R}^{m_1\times m_1\times m_3}$ *and* $V \in \mathbb{R}^{m_2\times m_2\times m_3}$ *hold.*

Proof. $\overline{U}^{(K)}$ and $\overline{V}^{(K)}$ are orthogonal matrices, because the singular value of the Matrix does not change with the multiplication of the orthogonal Matrix, and $\overline{X}^{(K)}$ and $U^{(K)}\overline{X}^{(K)}V^{(K)}$ have the same individual value; thus, $\|\mathrm{UXV}\| = \|X\|_{\mu}$ can be obtained. □

Proposition 3. $\|X\|_{\mu} \geq 0$ *holds for any tensor* $X \in \mathbb{R}^{m_1\times m_2\times m_3}$, $\|X\|_{\mu} = 0$ *if and only if* $X = 0$.

3.2.2. NLRTAPC Model. To eliminate the mixed noise of HSI, considering the global low-rank and local piecewise smooth characteristics of the band image, and combined with the PCTV regularization term in the nonconvex low-rank tensor approximation model, the mixed noise reduction method based on Nonconvex Low-Rank Tensor Approximation and Phase Consistency for Mixed Denoising (NLRTAPC) is presented by

$$\begin{aligned} &\arg\min_{\mathscr{L},\mathscr{I},\mathscr{G}} \|\mathbf{L}\|_{\mu} + \beta\|pc(\mathbf{Y})\|_{1,p} + \lambda\|\mathbf{I}\|_1 + \gamma\|\mathbf{G}\|_F^2, \\ &s.t. \quad \mathbf{L} + \mathbf{I} + \mathbf{G} = \mathbf{Y}, \end{aligned} \tag{8}$$

where β, λ, γ are nonnegative parameters, $\mathbf{L}$ is the restoring low-rank tensor, $\mathbf{I}$ is the sparse noise component, and $\mathbf{G}$ is

Table 1: Common nonconvex proxy functions.

Nonconvex proxy function	$G_\theta(x), x \geq 0, \theta > 0$	Supergradient $\nabla G_\theta(x)$
Logdet	$\theta/\log(\gamma+1)\log(\gamma x+1)$	$\gamma/(\gamma x+1)\log(\gamma+1)$
Gamma	$\theta x/x+\gamma$	$\theta\gamma/(x+\gamma)^2$
Lap	$\theta(1-\exp(-x/\gamma))$	$\theta/\gamma\exp(-x/\gamma)$
Ep	$\theta/1-\exp(-\gamma)(1-\exp(-\gamma x))$	$\theta\gamma/1-\exp(-)\exp(-\gamma x)$
Scad	$\begin{cases} \theta x, & x<0, \\ -x^2+2\gamma\theta x-\theta^2/2(\gamma+1)\theta, & x\leq\gamma\theta, \\ \theta^2(\gamma+1)/2, & x>\gamma\theta. \end{cases}$	$\begin{cases} \theta, & \text{if } x<\theta, \\ \gamma\theta-x/\gamma-1, & \lambda<x\leq\gamma\theta, \\ 0, & f\,x>\gamma\theta. \end{cases}$

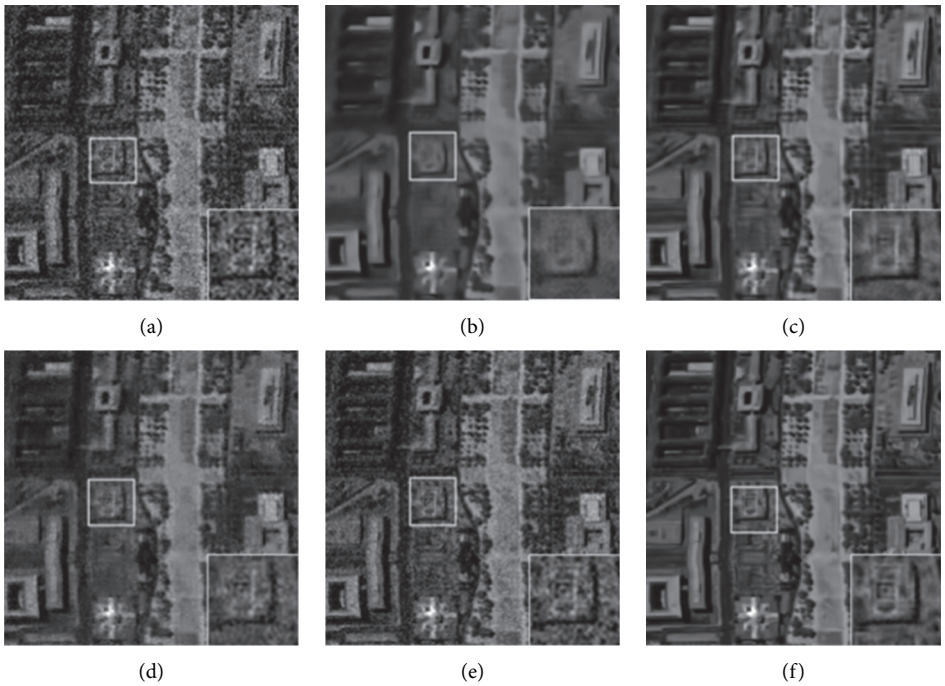


Figure 1: Comparison of Image restoration at 58 bands with noise type 4. (a) Noisy image. (b) BM4D. (c) NonLRMA. (d) GSSTV. (e) WGLRTD. (f) NLRTAPC.

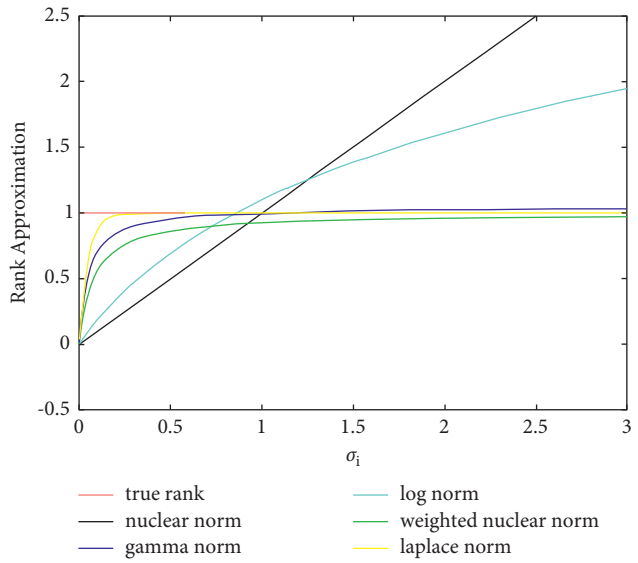


Figure 2: Approximating rank functions with different proxy functions.

the gauss noise component. The first term of the model is the low-rank tensor approximated by a nonconvex function with formula 2.1. The second term is the Schatten p norm constraint term based on phase congruency, and the last two are for noise error constraints. The NLRTAPC model not only preserves the low-rank features of the image more accurately than the tensor kernel norm model, but also uses the phase consistency to preserve the structural information such as edge and texture in the band restoration component.

3.3. Solution of NLRTAPC Model. NLRTAPC model is a nonconvex optimization problem, which is solved by decomposition subproblem. The ADMM algorithm can be used to solve the corresponding model iteratively. By introducing the secondary variable $\mathbf{Z}$, the NLRTAPC model can be expressed as

$$\begin{aligned} &\arg\min_{\mathscr{L},\mathscr{I},\mathscr{G},\mathscr{Z}} \|\mathbf{L}\|_{\mu} + \beta\|\mathbf{Z}\|_{1,p} + \lambda\|\mathbf{I}\|_1 + \gamma\|\mathbf{G}\|_F^2, \\ &\quad s.t. \quad \mathbf{L}+\mathbf{I}+\mathbf{G}=\mathbf{Y}, \mathbf{Z}=pc(\mathbf{Y}). \end{aligned} \tag{9}$$

Since the four variables $\mathbf{L}$, $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{Z}$ in the model are separable, the objective function can be solved by ADMM with formula 2.2, which is expressed as

$$\begin{aligned} M(\mathbf{L},\mathbf{I},\mathbf{G},\mathbf{Z};\mathbf{W},\mathbf{B}) &= \|\mathbf{L}\|_{\mu} + \beta\|\mathbf{Z}\|_{1,p} + \lambda\|\mathbf{I}\|_1 + \gamma\|\mathbf{G}\|_F^2 + \langle \mathbf{L}+\mathbf{I}+\mathbf{G}-\mathbf{Y}, \frac{\mathbf{W}}{\rho}\rangle \\ &+ \frac{\rho}{2}\left(\left\|\mathbf{L}+\mathbf{I}+\mathbf{G}-\mathbf{Y}+\frac{\mathbf{W}}{\rho}\right\|_F^2 + \langle \mathbf{Z}-pc(\mathbf{Y}), \frac{\mathbf{B}}{\rho}\rangle + \left\|\mathbf{Z}-pc(\mathbf{Y})+\frac{\mathbf{B}}{\rho}\right\|_F^2\right), \end{aligned} \tag{10}$$

where $\mathbf{W}$ and $\mathbf{B}$ are the Lagrange multiplier with $\mathbf{L}+\mathbf{I}+\mathbf{G}=\mathbf{Y}$ and $\mathbf{Z}=pc(\mathbf{L})$ constraints, respectively, and ρ is the penalty parameter. Iterating through the NLRTAPC model according to the ADMM framework can be broken down into the following five steps, each of which solves the corresponding variable, as follows:

(1) fix other tensor variables $\mathbf{I}$, $\mathbf{G}$, $\mathbf{Z}$, $\mathbf{W}$ and $\mathbf{B}$, and update the estimator $\mathbf{L}$. Under the iterative framework of $(k+1)$, the estimated true image is as follows:

$$\begin{aligned} \mathbf{L}_{k+1} &= \arg\min_{\mathbf{L}} M_{\rho}(\mathbf{L},\mathbf{I}_k,\mathbf{G}_k,\mathbf{Z}_k;\mathbf{W}_k,\mathbf{B}_k) \\ &= \arg\min_{\mathbf{L}} \|\mathbf{L}\|_{\mu} + \frac{1}{\rho_k}\|\mathbf{L}-\mathbf{D}_k\|_F^2. \end{aligned} \tag{11}$$

where $\mathbf{D}_k = 1/2(\mathbf{Y}-\mathbf{I}_k-\mathbf{G}_k-\mathbf{Z}_k)-\mathbf{W}_k/\rho_k$.

According to Proposition 2, given any given tensor $\mathbf{Z}$, it can be decomposed into $\mathbf{Z}=U*S*V^T$ by T-SVD, and $\arg\min_{\mathbf{Y}} \|\mathbf{Y}\|_{\mu} + \lambda/2\|\mathbf{Y}-\mathbf{Z}\|_F^2$ can be solved by a weighted tensor singular threshold method.

The tensor rank approximation problem of nonconvex normalized μ is used. The problem can be solved by the singular threshold operator of tensor t-SVD decomposition. When $\mathbf{D}_k$ is decomposed into $\mathbf{D}_k=\mathbf{U}*\mathbf{S}*\mathbf{V}^T$ by T-SVD, the optimal solution is $\mathbf{L}_{k+1}=\mathbf{U}*\mathbf{D}_{\nabla g(\sigma)/\rho_k}*\mathbf{V}^T$, where $\mathbf{D}_{\nabla g(\sigma)/\rho_k}$ is the diagonal tensor of f, and the elements of the diagonal tensor are obtained by Fourier domain calculation; that is, $\ⅅ_{\nabla g(\sigma)/\rho_k}{}^{(i)}(i,k) = max((S^{(i)}(i,k)-\nabla g(\sigma^{i,k})/\rho_k), 0)$.

(2) fix the other tensor variables $\mathbf{D}$, $\mathbf{I}$, $\mathbf{G}$, $\mathbf{W}$ and $\mathbf{B}$, update the l_p norm constraint $\mathbf{Z}$ of PC, and save the spatial smoothness and edge information.

$$\mathbf{Z}_{k+1} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{Z}\|_{1,p} + \frac{1}{\rho_k}\|\mathbf{Z}-\mathbf{T}_k\|_F^2. \tag{12}$$

where $\mathbf{T}_k = pc(\mathbf{Y}_k) + \mathbf{B}_k/\rho_k$. First, we use singular value decomposition tensor $\mathbf{T}_k$ with T-SVD, $\mathbf{T}_k=\mathbf{U}_k^1*\mathbf{S}_k^1*\mathbf{V}_k^{1T}$. Then, lp norm approximation is performed for the f diagonal tensor $\mathbf{S}_k^1$, denoted as $\text{prox}_{\rho_k/2}(\sigma(\mathbf{S}_k^1))$. Finally, $\mathbf{Z}_{k+1}$ is reconstructed, and $\mathbf{Z}_{k+1}=\mathbf{U}_k^1*\text{dialog}(\text{prox}_{\rho_k/2}(\sigma(\mathbf{S}_k^1)))*\mathbf{V}_k^{1T}$, where dialog $(\cdot)$ means that the vector elements are expressed in a matrix form, is a diagonal matrix convenient operation.

(3) To remove impulse noise, other tensor variables $\mathbf{L}$, $\mathbf{G}$, $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{Z}$ are fixed, and $\mathbf{I}$ is updated.

$$\begin{aligned} \mathbf{I}_{k+1} &= \arg\min_{\mathscr{I}} \frac{\lambda}{\rho_k}\|\mathbf{I}\|_1 + \frac{1}{2}\|\mathbf{I}-\mathbf{Q}_k\|_F^2 \\ &= \sum_{i=1}^{B} \arg\min_{I^{(i)}} \frac{\lambda}{\rho_k}\|I^{(i)}\|_1 + \frac{1}{2}\|I^{(i)}-Q_k^{(i)}\|_F^2. \end{aligned} \tag{13}$$

where $Q_k^{(i)} = Y^{(i)} - L_{k+1}^{(i)} - G_k^{(i)} - W_k^{(i)}/\rho_k$, and then the closed-form solution $I_{k+1} = \text{sign}(\mathbf{Q}_k)\max\{|\mathbf{Q}_k| - \lambda/\rho_k, 0\}$ of formula (13) is obtained by using the contraction operator of Matrix elements.

(4) To remove gauss noise, fix $\mathbf{L}$, $\mathbf{I}$, $\mathbf{W}$, $\mathbf{B}$, and $\mathbf{Z}$, and update $\mathbf{G}$.

$$\begin{aligned} \mathbf{G}_{k+1} &= \arg\min_{\mathbf{G}} \gamma\|\mathbf{G}\|_F^2 + \frac{\rho_k}{2}\|\mathbf{G}-\mathbf{R}_k\|_F^2 \\ &= \sum_{i=1}^{B} \arg\min_{G^{(i)}} \gamma\|G^{(i)}\|_F^2 + \frac{\rho_k}{2}\|G^{(i)}-R_k^{(i)}\|_F^2. \end{aligned} \tag{14}$$

where $R_{k+1}^{(i)} = Y_k^{(i)} - L_{k+1}^{(i)} - I_{k+1}^{(i)} - W_k^{(i)}/\rho_k$, and $\mathbf{W}_k$ is the Lagrange multiplier of the k-th iteration.

Formula (14) is a standard least square regression problem, which is easy to solve.

(5) Update Lagrange multiplier and penalty parameters.

$$\begin{cases} \mathbf{W}_{k+1} = \mathbf{W}_k + \rho_k\left(\mathbf{L}_{k+1} + \mathbf{G}_{k+1} + \mathbf{I}_{k+1} - \mathbf{Y}\right), \\ \mathbf{B}_{k+1} = \mathbf{B}_k + \rho_k\left(\mathbf{Z}_{k+1} - pc\left(\mathbf{Y}_k\right)\right), \\ \rho_{k+1} = \min\left\{\kappa \times \rho_k, \rho_{\max}\right\}. \end{cases} \tag{15}$$

where κ is the contraction parameter; set $\kappa > 1$ to accelerate the rate of convergence and $\rho_{\max}$ to be the maximum value of ρ.

The NLRTAPC model is described by ADMM in Algorithm 1. To obtain good phase consistency, the front slice of the HSI image is a waveband image, and the phase consistency is extracted in Matrix form to get pc (y).

4. Experimental Results and Analysis

In this section, experiments were carried out to demonstrate the mixed noise removal capability of our model. To better illustrate the superiority of the combination of the nonconvex smooth rank approximation model and the l_p norm constraint, the validity of the proposed method is verified by simulation and real experiments, and the quantitative and visual performance of the four advanced HSI denoising methods are compared with the denoising results of this method. These methods include block-matched 4D filtering (BM4D), Nonlrma [14], global spatial something spectral total variation (GSSTV) [15], and weighted group sparse regularized low-rank tensor decomposition (WGLRTD). The code of all comparison methods is Matlab Code. All experiments of comparison method are carried out on Intel Core i7-4970 CPU (3.60 GHz) and 16 GB RAM computer using Matlab R2018a. The parameters of these methods in the experiment are set according to the method suggestions to obtain the best performance.

4.1. Simulation Data Experiment. Two data sets, namely, WDC and Indian data sets, are used in the simulation data experiment. The WDC data set was collected by a hyperspectral digital image acquisition experiment (HYDICE) sensor at a mall in Washington, D. C., at 191 wavelengths. The "Indian" is a collection of $145 * 145$ pixels and 224 spectral reflectivity bands collected by the AVIRIS sensors over the Indian pine proving ground in northwestern Indiana. Since the data set contains a portion of atmospheric absorption bands that are not useful for subsequent applications, the 200-band noise-free images were selected for the experiment. Experiment parameters are set by $\beta = 0.1, \lambda = 1, \gamma = 0.001, \mu = 0.02, \rho = 0.1, \kappa = 1.02$.

To simulate the complex noise situation in the real scene, Matlab is used to generate 4 kinds of mixed noise in the two clean HSI data sets. For noise type 1 to noise type 9, the intensity of gauss noise, random noise, and band noise is shown in Table 2.

4.2. Analysis of Simulation Experiment Results. There are four methods in the experiment, BM4D, NonLRMA, GSSTV, and WGLRTD. Three noise types (noise types 1, 4, and 7) are chosen randomly to compare the performance of each algorithm model. Figure 1 shows the result of a 58 band data recovery from a mall in Washington, D. C., under noise Type 4 conditions. For better visual contrast, some areas of the image are deliberately magnified, as shown in Figure 1:

As can be seen from Figure 1, BM4D has a serious mold and loss of image detail due to oversmoothing. NonLRMA and GSSTV still retain a small amount of noise, and the image is darkened due to the offset of the whole pixel value. In the enlarged area, the WGLRTD left a lot of random noise, and the details of the image were not well preserved. On the contrary, the NLRTAPC model can eliminate all the mixed noises and preserve the edges and details of the image effectively, which shows that the NLRTAPC method has better performance in recovering WDC data set than other current typical methods.

Figure 3 shows a comparison of restoration at the 165 bands of the Indian dataset under noise type 1, from the overall view of Figure 3; since the damage to the image structure caused by noise type 1 is not very serious, several methods can ensure that the overall structure recovers well, but it can be found that NLRTAPC, the Algorithm model designed in this chapter, has a clear structure in the edge region, while it still keeps the smooth effect in the smooth region and presents a better visual effect.

In Figure 4, the 165 band data recovery results are for the Indian Dataset. By comparing and analyzing the recovery results of all the comparison methods, a more direct-viewing result is obtained. A more direct result can be obtained from Figure 4. Because the original image has been destroyed seriously, the recovery result of the GSSTV method is ambiguous, the recovery result of the NonLRMA method still has an "artificial gradient" phenomenon, and bM4D and WGLRTD methods have some modulus on the edge of the image structure. In this chapter, HLRTD-SSTV Algorithm is presented to restore the result of the overall structure clear and smooth, and the visual effect is good; overall, it is superior to other methods. To demonstrate the excellent performance of the NLRTAPC algorithm model, two objective quantitative evaluation indexes are introduced to prove the performance of the NLRTAPC algorithm model in all simulation experiments, which are average Peak signal-to-noise ratios (MPSNR) of all bands and average structural similarity (MSSIM) of all bands.

Tables 3 and 4 show the MPSNR and MSSIM values of the Indian and WDC datasets under various mixed noise types, respectively. In the experiment, MPSNR and MSSIM were evaluated as the average of all bands of the data set, and NLRTAPC gets good indicator values. From the evaluation data of nine different noise types, the NLRTAPC model achieves the best performance in most cases. With the increase of noise level, the performance degradation is relatively slow compared with other methods. In particular, the MSSIM Algorithm can improve the performance obviously, and the key is that the phase consistency constraint is used to capture the edge information of the image more comprehensively and more accurately than TV.

input: Observational data $\mathbf{Y}$, Index set Ω, Parameters $\beta > 0, \lambda > 0, \gamma > 0$
Initializing: $\mathbf{L} = \mathbf{Y}$, $\mathbf{I} = 0, \mathbf{G} = 0, \mathbf{Z} = pc(\mathbf{Y}), \mathbf{W} = 0, \mathbf{B} = 0$, $k = 0$, num = 1000 and $\varepsilon = 10^{-5}, i = 0$.
(1) when $\|L^{l+1} - L^l\|_F / \|L^l\|_F \geq \varepsilon$ or $k <$ num.
(2) k ++;
(3) $\overline{Y}^k = fft(\mathbf{Y}^{(k)}, , 3)$;
(4) Cycle ($I = 1$) until the maximum band number B of the observed data
(5) The updated estimate $L^{(i)}$ is calculated according to formula (11);
(6) The updated estimate $Z^{(i)}$ is calculated according to formula (12);
(7) The updated estimate $I^{(i)}$ is calculated according to formula (13);
(8) The updated estimate $G^{(i)}$ is calculated according to formula (14);
(9) According to formula (15), the Lagrange multiplier and the penalty parameters are updated respectively.
(10) End;
(11) Using fold(unfold(•)) to restore the tensor structure $\mathbf{L}$, $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{Z}$, $\mathbf{Y}^{(k+1)} = \mathbf{L}^{(k+1)}$;
(12) $Y^{K+1} = iffft(\mathcal{Y}^{k+1}, , 3)$;
End;
output: denoising tensor $\mathbf{L}^{(k+1)}$.

Algorithm 1: Solve the proposed NLRTAPC model with ADMM

Table 2: Description of noise type.

Noise type	Gaussian noise	Random noise	Band noise
Type 1	Mean:0 Noise variance:10	10% impulse random noise	Band:50–60 Stripe width:2 pixels
Type 2	Mean:0 Noise variance:20	10% impulse random noise	Band:50–60 Stripe width:2 pixels
Type 3	Mean:0 Noise variance:30	10% impulse random noise	Band:50–60 Stripe width:2 pixels
Type 4	Mean:0 Noise variance:10	20% impulse random noise	Band:50–60 Stripe width:2 pixels
Type 5	Mean:0 Noise variance:20	20% impulse random noise	Band:50–60 Stripe width:2 pixels
Type 6	Mean:0 Noise variance:30	20% impulse random noise	Band:50–60 Stripe width:2 pixels
Type 7	Mean:0 Noise variance:10	30% impulse random noise	Band:50–60 Stripe width:2 pixels
Type 8	Mean:0 Noise variance:20	30% impulse random noise	Band:50–60 Stripe width:2 pixels
Type 9	Mean:0 Noise variance:30	30% impulse random noise	Band:50–60 Stripe width:2 pixels

4.3. Real Dataset Experiment. In the simulation experiment, the performance of the NLRTAPC under various noise conditions is evaluated by visual, quantitative, and qualitative methods. This section performs a hyperspectral digital image acquisition experiment (HYDICE) on a real urban dataset to verify the effectiveness of NLRTAPC's real hyperspectral image restoration. The whole dataset is applied in the experiment of denoising algorithm. Some brands of the dataset are heavily polluted by the atmosphere, hygroscopicity, and some thermal noise; in addition, some bands are polluted by striations, gauss noise, and random noise. As a preprocessing, the pixel values for each band are normalized to [0,1].

Figure 5 shows the results of band 76 recoveries in this data. As can be seen from the graph, the original band image is corrupted by various noises, including gauss noise and unknown noise. After using different HSI reconstruction methods, the noise was removed, BM4D mode was burnt, and the WGLRTD method still had some noise in the denoising band image. From the enlarged area and the whole image, it can be seen that the NLRTAPC Algorithm in this chapter is effective to remove the mixed noise while preserving the image structure. Figure 6 shows the results of band 136 recoveries in this data. In contrast, the NLRTAPC approach can completely suppress all kinds of noise, effectively preserving details. Compared with other methods, the image restored by this method is smoother. These visualization results show that the NLRTAPC method can remove more complex noises hidden in HSI than other methods.

To demonstrate the results of image reconstruction, Figure 7 uses different methods to estimate spectral characteristic curves for 76 band images in real-world urban datasets. In Figure 7, the ordinal of bands is shown on the horizontal axis, and the average of the estimated spectral

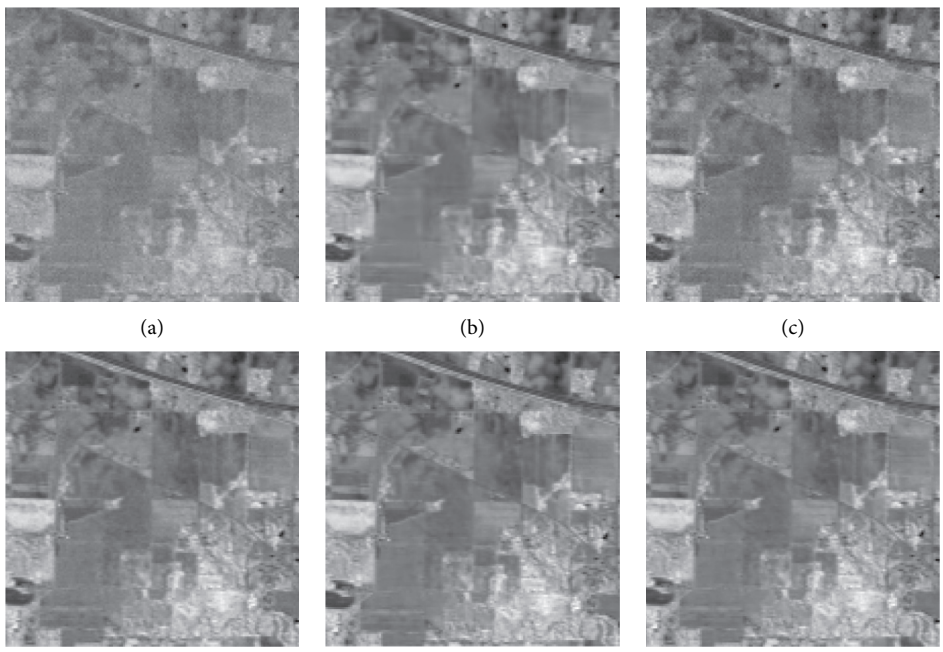


Figure 3: Comparison of Image restoration at 165 bands with noise type 1. (a) Noisy image. (b) BM4D. (c) NonLRMA. (d) GSSTV. (d) WGLRTD. (f) NLRTAPC.

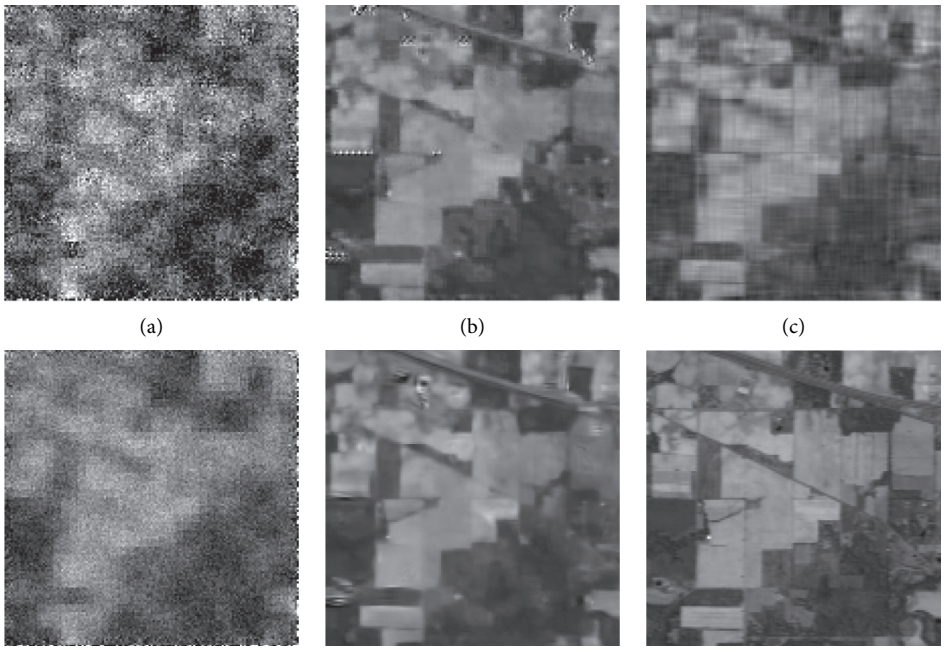


Figure 4: Comparison of Image restoration at 165 band with noise type 7. (a) Noisy image. (b) BM4D. (c) NonLRMA. (d) GSSTV. (e) WGLRTD. (f) NLRTAPC.

characteristics of each band is shown on the vertical axis. As shown in Figure 5(a), there are many fluctuations in the spectral characteristic curve of the original band image due to striations and other noises. It can be observed that the comparison method suppressed the volatility to some extent after the recovery of the different methods shown in Figures 5(b)–5(f). The results of spectral characteristic curves show that the NLRTAPC recovery curve is more smooth than other methods, which indicates that NLRTAPC has a better denoising effect. As shown in Figures 7(b) and

Table 3: MPSNR and MSSIM values of Indian dataset under mixed noise types.

Noise type	BM4D	NonLRMA	GSSTV	WGLRTD	NLRTAPC
Type 1	35.29/0.9628	35.30/0.9541	35.74/0.9386	35.49/0.9309	36.40/0.9701
Type 2	33.31/0.9110	34.10/0.9350	33.50/0.9214	33.28/0.9154	35.20/0.9521
Type 3	28.10/0.8630	30.90/0.8769	29.09/0.8795	29.05/0.8580	30.95/0.8798
Type 4	34.30/0.9326	34.25/0.9381	34.51/0.9379	34.43/0.9239	35.61/0.9481
Type 5	32.57/0.9089	32.12/0.9101	32.29/0.9091	32.16/0.9030	32.59/0.9162
Type 6	28.21/0.8689	30.41/0.8990	30.13/0.8921	30.05/0.8849	29.99/0.8879
Type 7	28.12/0.8541	28.50/0.8979	28.19/0.8897	28.10/0.8789	29.98/0.9169
Type 8	26.39/0.9089	26.53/0.9101	26.59/0.9119	26.40/0.9081	26.59/0.9131
Type 9	24.73/0.8529	24.86/0.8515	24.81/0.8430	24.69/0.8351	24.99/0.8579

Table 4: MPSNR and MSSIM metric values of WDC dataset under mixed noise types.

Noise type	BM4D	NonLRMA	GSSTV	WGLRTD	NLRTAPC
Type 1	36.23/0.9653	36.29/0.9680	36.64/0.9689	36.64/0.9639	36.95/0.9801
Type 2	34.67/0.9220	35.76/0.9469	35.95/0.9508	35.84/0.9500	36.03/0.9527
Type 3	30.31/0.8970	31.87/0.9171	32.13/0.9194	32.05/0.9166	32.26/0.9210
Type 4	35.31/0.9190	35.80/0.9349	35.90/0.9419	35.79/0.9411	36.10/0.9521
Type 5	32.89/0.9081	33.61/0.9221	33.81/0.9259	33.78/0.9250	33.98/0.9289
Type 6	29.42/0.8740	30.41/0.9094	30.81/0.9102	30.51/0.9102	31.08/0.9126
Type 7	32.51/0.9082	33.41/0.9279	33.65/0.9296	33.52/0.9271	33.97/0.9319
Type 8	29.61/0.9094	30.36/0.9118	30.90/0.9154	30.71/0.9139	31.25/0.9189
Type 9	27.81/0.8677	28.80/0.8862	28.94/0.8893	28.81/0.8861	29.21/0.8980

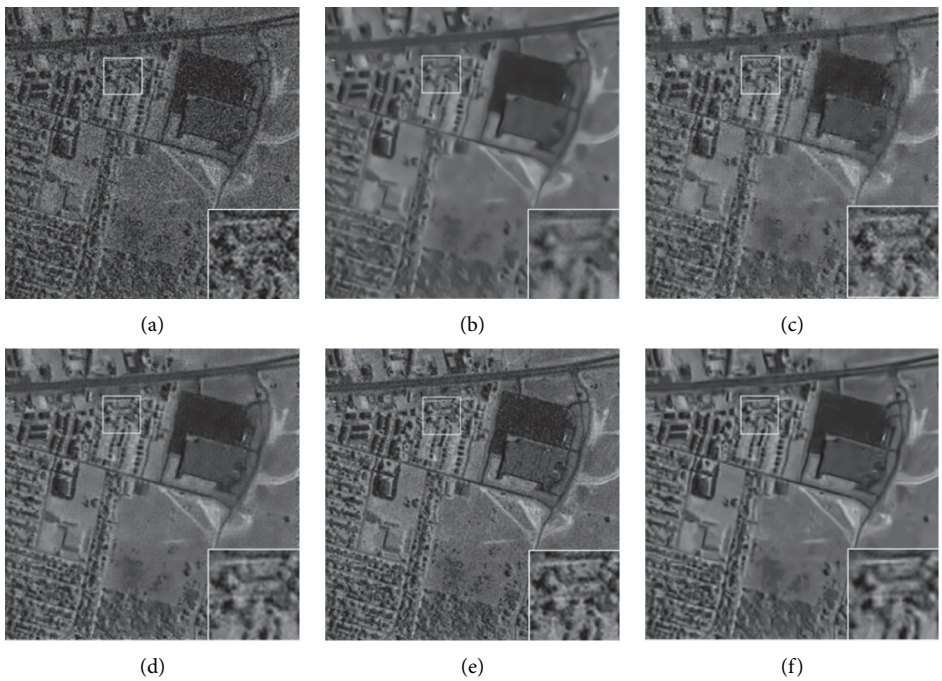


Figure 5: The effect comparison chart of recovering the 76th band image of Hydice Dataset with 5 algorithms. (a) Real image. (b) BM4D. (c) NonLRMA. (d) GSSTV. (e) WGLRTD. (f) NLRTAPC.

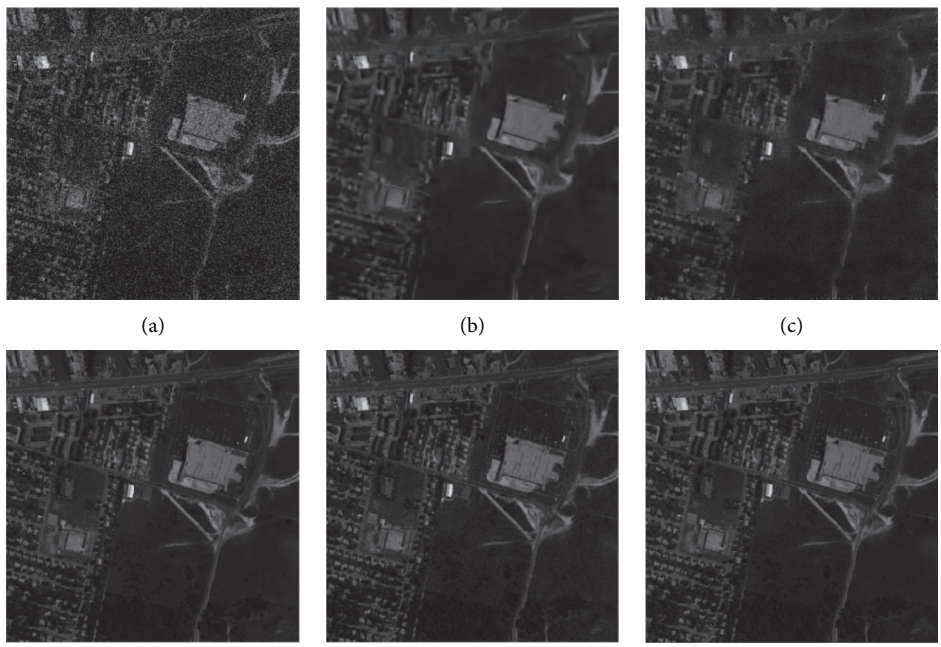


Figure 6: The effect comparison chart of recovering the 136th band image of Hydice Dataset with 5 algorithms.

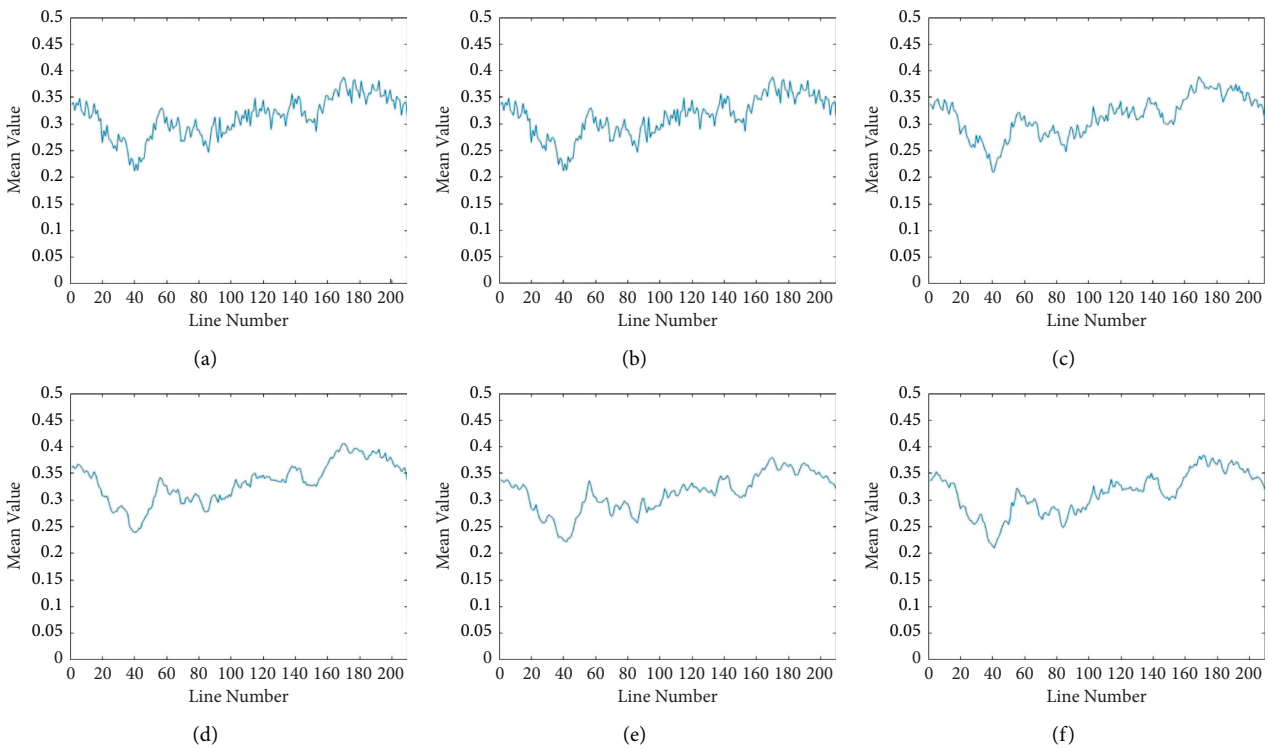


Figure 7: The spectral characteristic curves of 76 band images in Hydice Dataset are estimated by different methods. (a) Real image. (b) BM4D. (c) NonLRMA. (d) GSSTV. (e) WGLRTD. (f) NLRTAPC.

7(c), there are also small fluctuations in the curve, indicating that some of the mixed noise in the image remains in the image.

5. Conclusion

In this paper, a hybrid noise removal method based on nonconvex low-rank tensor approximation is proposed, based on the optimization method ADMM proposed to solve the model by using the structural information of all bands and the local information of the neighborhood. NLRTAPC model can make use of the structure information of all bands and local neighborhood information more effectively, promote the removal of mixed noise, and greatly alleviate the "artificial ladder" phenomenon. The results of simulation and real data experiments show that the proposed method is more effective than the competing state-of-the-art denoising methods. The future improved direction of this approach is to optimize parameters of NLRTAPC to improve model adaptability.

Data Availability

The data used to support the findings of this study were supplied by Li bo under license and so cannot be made freely available. Requests for access to these data should be made to Li Bo (sctulibo2021@126.com).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was jointly funded by the Sichuan Science and Technology Program, Grant/Award Numbers: 2019ZYZF0169; the Science and Technology Support Project of Panzhihua City of Sichuan Province of China under Grant 2021CY-S-6, and Tourism resources big data application technology innovation team project.

References

[1] D. Letexier and S. Bourennane, "Multidimensional wiener filtering using fourth order statistics of hyperspectral images," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pp. 917–920, IEEE, Las Vegas, NV, USA, March 2008.

[2] G. Chen, T. D. Bui, and A. Krzyzak, "Denoising of three-dimensional data cube using bivariate wavelet shrinking," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 3, pp. 403–413, 2011.

[3] H. Hongyan Zhang, W. Wei He, L. Liangpei Zhang, H. Huanfeng Shen, and Q. Qiangqiang Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2014.

[4] R. Zhu, M. Dong, and J.-H. Xue, "Spectral nonlocal restoration of hyperspectral images with low-rank property," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3062–3067, 2015.

[5] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the 2014 IEEE Conference on IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 2862–2869, Columbus, OH, USA, June 2014.

[6] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Hyperspectral image recovery employing a multidimensional nonlocal total variation model," *Signal Processing*, vol. 111, pp. 230–248, 2015.

[7] J. Jiang, J. Yang, Y. Cui, and L. Luo, "Mixed noise removal by weighted low rank model," *Neurocomputing*, vol. 151, no. 2, pp. 817–826, 2015.

[8] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: a low-rank approach," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 700–711, 2013.

[9] Y.-M. Huang, H.-Y. Yan, Y.-W. Wen, and X. Yang, "Rank minimization with applications to image noise removal," *Information Sciences*, vol. 429, no. 6, pp. 147–163, 2018.

[10] Z. Wu, Q. Wang, J. Jin, and Y. Shen, "Structure tensor total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising," *Signal Processing*, vol. 131, no. 1, pp. 202–219, 2017.

[11] T. Huang, W. Dong, X. Xie, G. Shi, and X. Bai, "Mixed noise removal via laplacian scale mixture modeling and nonlocal low-rank approximation," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3171–3186, 2017.

[12] M. C. Morrone and R. A. Owens, "Feature detection from local energy," *Pattern Recognition*, vol. 6, no. 5, pp. 51-52, 1994.

[13] X.-G. Luo, H.-J. Wang, and S. Wang, "Monogenic signal theory based feature similarity index for image quality assessment," *AEU - International Journal of Electronics and Communications*, vol. 69, no. 1, pp. 75–81, 2015.

[14] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5366–5380, 2017.

[15] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial-spectral total variation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 713–729, 2018.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 1312058, 10 pages
https://doi.org/10.1155/2021/1312058

Research Article

Optimization of Vehicle Transportation Route Based on IoT

Qian Yu, Yuanguo Wang, Xiaogang Jiang, Bailu Zhao, Xiuling Zhang, Xiaobei Wang, and Qingqing Liu

College of Information Engineering, Binzhou Polytechnic, Binzhou 256603, Shandong, China

Correspondence should be addressed to Xiaogang Jiang; jxg@bzpt.edu.cn

Received 13 July 2021; Revised 7 September 2021; Accepted 13 September 2021; Published 30 September 2021

Academic Editor: Xianyong Li

With the rapid development of logistics industry, optimization of road transport has become a constraint that must be overcome in the development of related industries. In the IoT era, classic car routing solutions could not meet many different needs. The relevant research findings are endless but not suitable to reduce costs in logistics and distribution processes and meet the needs of customers. This paper researches on vehicle path optimization using IoT technology and intelligent algorithms. Firstly, the traditional GA is optimized, and its coding mode, fitness function, selection, crossover, and mutation operators are studied. The crossover probability was set to 0.6, and the mutation probability was set to 0.1; then, according to the improved GA, a vehicle route optimization model was created. Finally, simulations were conducted to optimize vehicle routes for some distribution centers and 15 customer sites, and the model's validity was tested. Experimental data show that the improved genetic algorithm begins to converge in 100 generations with a running time of 37.265 s. We calculate the time sensitivity of the customer. An algorithmic model is then used to determine distribution plans based on product demand and time sensitivity. In addition, we compare distribution costs and customer satisfaction of algorithmic and randomized plans. The distribution cost and customer satisfaction of the algorithmic and random patterns were 498.09 yuan and 573.13 yuan and 140.45 and 131.35, respectively. This shows that the vehicle routing optimization model using IoT technology and an improved GA can reduce distribution costs and increase customer satisfaction.

1. Introduction

1.1. Background Significance. The IoT is an important carrier for collecting, transmitting, processing, and applying information. Related technologies such as big data, cloud computing, sensors, and so on are widely used in various fields, especially the logistics industry [1]. The continuous improvement of the urban system makes the establishment of a good urban distribution system, a problem that must be solved to increase the development of the urban economic system. The optimization of the transportation path of the distribution vehicle is a vital link in the urban logistics distribution system, which connects the production line, warehouses, and consumers [2].

1.2. Related Work. IoT technology is widely used in many fields. Based on the latest research on smart homes and IoT technology, Hui et al. outlined specific requirements for smart home construction and proposed requirements based on specific quality specifications of smart homes. The building blocks are divided into seven independent sections [3]. Wu offered the design and management of museum collections and RFID-based intelligent navigation systems, a new way to build smart museums [4]. It gave everyone a new understanding of the IoT, but their experimental sources were not clear enough. As a result, the research findings do not have exact references. Optimization of vehicle routes has always been a key issue in logistics [5]. Discussed by local and international researchers, Braekers conducted a categorization review of the literature on vehicle routing issues published between 2009 and June 2015, based on a modified version of the taxonomy at http://www.braekers.com. The 277 existing coverages were classified, and the development trend of VRP literatures was analyzed [6]. Yao et al. proposed the box-to-collection station heterogeneous vehicle routing problem and used particle swarm optimization (PSO) to solve the problem [7]. To improve the efficiency of

the particle swarm algorithm, he adopted adaptive inertia weights and a local search strategy. Their research data are quite old, which are quite different from the actual data under current research background.

1.3. Innovative Points in This Paper. The research innovations are as follows: (1) to study the coding mechanisms of GAs: fitness services, selection, crossover, and conversion operators, and improve GA by scheduling relevant events; (2) based on advanced GA and the problem of optimizing vehicle transportation, a GA path optimization model was created; and (3) the model outlined in this paper was used to perform mock experiments to test the algorithm's effectiveness, and the distribution cost and customer satisfaction of distribution schemes under an algorithmic model and random distribution model were compared. The results show that the algorithm model proposed in this paper can increase customer satisfaction and save distribution cost.

2. IoT Technology and Optimization Algorithm for Vehicle Transportation Path

2.1. Key Technologies of the IoT

2.1.1. GPS Positioning and Wireless Sensor Network. There are 24 GPS satellites evenly distributed over the surface [8], and navigation information can be stored on the satellites [9].

The sensor node performs a preliminary fusion of the detected data and sends the data and its own position information to the sink node in the form of multihop transmission, and finally the detected data are sent to the end user via satellite and Internet.

The positioning method of the wireless sensor network is to equip a very small number of nodes with GPS, relying on the information interaction between the nodes, and using the positioning algorithm to calculate the location. This positioning process includes two stages: distance estimation and coordinate calculation. The calculation of the algorithm performance generally starts from the normal position and the weight of the issue. Among them, the calculation method of positioning accuracy is shown in the following formula.

$$\text{Error} = \frac{\sqrt{(x_i - p_i)^2 + (y_i - q_i)^2}}{r}. \quad (1)$$

2.1.2. Data Fusion Method. Kalman filtering is an unbiased minimum method. The statistical properties of the measurement model are used to recursively determine the best fusion data estimate in the statistical sense. The Kalman filter uses the state space of a linear random system. In operation, the Kalman filter method uses the minimum variance of the linear deviation as the standard to change the state parameters of the filter [10].

Bayesian estimation is used for multiple sensor information fusion. We use the conditional probabilistic formula for processing and finally give a systematic decision based on several rules. Bayesian estimation is an effective method to fuse information from multiple sensors in a static environment.

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}. \quad (2)$$

Fuzzy theory is based on the basis of fuzzy sets, and the uncertain fuzzy number can be recognized by the computer after transformation. The received information is processed by fuzzy logic, and a rule basis is established based on experience. Finally, we compare the current state of sensor monitoring with the basic state to determine the reliability of the information. If the information reliability is high, we use fuzzy variables to represent the actual variables and obtain the actual variable values by solving the fuzzy variables [11]. This method uses a systematic method to model the uncertainty of the fusion process and finally produces consistent fuzzy inferences. This approach can solve the problem of information conflict and inconsistency to some extent, but it requires the work of a confusing member, so the ability to learn and adapt is not enough.

2.1.3. Cloud Computing and Data Mining. Cloud computing is a new technology. At present, some large Internet companies have established super-servers to provide storage and computing data centers. The cloud database brings great convenience to small- and medium-sized enterprises and individual users. Cloud computing is an Internet-based computing model that the public participates in. Its computing resources (computing functions, storage functions, and interactive functions) are dynamic, scalable, and virtualized and can be provided as services [12].

The foundation of data mining is data warehouse. Data mining is the use of related mining tools to classify potential connections between data according to the needs of database information, thereby promoting the popularization and utilization of information. Finding understanding is a product of an organic combination of artificial intelligence and database, which enables it to understand native languages and have the ability to speak.

2.2. Optimization of Vehicle Transportation Path

2.2.1. Elements of the Problem of Vehicle Transportation Routes. The vehicle routing problem refers to the use of vehicles to provide delivery services to customers with various needs, optimize the overall distribution channels of the organization, maximize the satisfaction of the needs of each customer, and achieve the corresponding goals under the constraints of time and load [13]. The constituent elements of the vehicle routing problem include distribution centers, customers, vehicles, distribution network, parking lots, constraints, and goals [11]. Commonly used constraints include vehicle load constraints, travel distance constraints, and time window constraints. The usual goals to be achieved include the lowest total cost of delivery, the shortest delivery distance, and the shortest delivery time.

2.2.2. Mathematical Model of Vehicle Transportation Route Problem. The general transportation problem of automobiles is to make the total transportation distance the shortest or the transportation cost the lowest [14].

Define graph $G = (S, L), S = \{s_1, s_2, \ldots, s_n\}$, distribution route $L = \{(s_i, s_j), s_i, s_j \in S\}$, distribution center s_0, number of vehicles M, vehicle load G of each vehicle, customer point $s_i (i = 1, 2, \ldots, n)$, customer point demand $(g_1, g_2, \ldots, g_n)$, and cost f_{ij}. The mathematical model established is as follows:

$$x_{ijm} = \begin{cases} 1, \text{vehicle } m \text{ travels from point } i \text{ to point } j, \\ 0, \text{otherwise}, \end{cases} \left(i, j = 0, 1, \ldots, n; ; m = 1, 2, \ldots, M\right), \tag{3}$$

$$x_{im} = \begin{cases} 1, \text{customer point } i \text{ is served by vehicle } m, \\ 0, \text{otherwise}, \end{cases} \left(i, j = 0, 1, \ldots, n; ; m = 1, 2, \ldots, M\right). \tag{4}$$

Assuming that the delivery vehicles are uniform and the delivery task is a one-way task, only the delivery problem needs to be considered. Formulas (3) and (4) are two decision variables. The goal is the lowest total transportation cost, and the optimization objective function is shown in the following formula:

$$\min Z = \sum_i \sum_j \sum_m f_{ij} x_{ijm}, \tag{5}$$

$$\sum_i g_i y_{im} \le G, \quad (m = 1, 2, \ldots, M), \tag{6}$$

$$\sum_m y_{im} = 1, \quad (i = 1, 2, \ldots, n), \tag{7}$$

$$\sum_m y_{s_0 m} = K, \tag{8}$$

$$\sum_i x_{ijm} = y_{mj}, \quad (j = 1, 2, \ldots, n), \tag{9}$$

$$\sum_j x_{ijm} = y_{mi}, \quad (i = 1, 2, \ldots, n). \tag{10}$$

Among them, formula (6) indicates that the cargo loaded by the vehicle cannot exceed the load of the vehicle. Formula (7) indicates that each customer point is on the delivery route. Formula (8) indicates that the distribution center of each vehicle is consistent. Formulas (9) and (10) indicate that if the customer points i, j need to be on the delivery route of the vehicle m, the service of the vehicle can be obtained [15].

2.2.3. Classification of Vehicle Transportation Path Problems. The important feature of the static vehicle routing problem is that the variables related to the vehicle are known before the vehicle scheduling is arranged and will not change after the vehicle scheduling [16].

According to the number of optimized objective functions, it is divided into single-objective problem and multiobjective problem. According to the demand characteristics, it can be divided into pure delivery, pure pickup, and integrated delivery and pickup vehicle routing problems. According to the vehicle type, it can be divided into single vehicle type and multiple vehicle type vehicle routing problems [17].

2.3. Intelligent Algorithm for Path Optimization

2.3.1. GA (Genetic Algorithm). GA can effectively solve the optimization problem of vehicle path. Even if the problem has discontinuities and nonlinearities or the internal structure of the problem is not clear, GAs can also solve these complex problems that cannot be mathematically modeled. The effective multibranch random parallel search feature makes the GA not easy to fall into the local optimum [18]. By increasing the number of populations, the optimization branches can be increased, which is very convenient for handling large-scale optimization problems, and the speed will also be faster.

First, change the optimal variables of the optimization problem to be optimized, then establish a function for individual compatibility calculation based on the function objectivity of the optimization problem. And randomly develop a group of individuals as the first solution point. If the conditions are not met, copy, cross, and mutate operations will be performed to verify again whether the termination conditions are met. Repeat the operation until the optimal solution of the degree of adaptation is obtained [19].

The fitness function is the optimization decision variable of the problem, which needs to be directly converted from the objective function of the mathematical model of the optimization problem. Its expression is shown in the following formula:

$$\zeta(x) = \text{Fit}(f(x)). \tag{11}$$

Considering that the probability requirement in the operation of GA is nonnegative, the aforementioned fitness function can be transformed into the following form:

$$\zeta(x) = \begin{cases} A_{\max} - f(x), & \text{if } f(x) < A_{\max}, \\ 0, & \text{else}, \end{cases} \tag{12}$$

$$\zeta(x) = \frac{1}{1 + A_{\max} + f(x)}, \tag{13}$$

where $A_{\max} + f(x) \geq 0$.

2.3.2. Ant Colony Algorithm. The solution to the problem of optimizing the vehicle path is considered the set of feasible paths for ants to find food. The longer the algorithm runs, the higher the pheromone concentration of the shorter path, and the solution to the problem will gradually converge [20]. Therefore, the highest and shortest paths of the pheromone are considered the optimal solution of the optimization problem.

The basic steps of the ant colony algorithm are as follows: first, initialize the parameters and put $x = (1, 2, \ldots, k, \ \ldots, X)$ ants in $y(y = 1, 2, \ldots, \ i, \ldots, \ j, \ldots, Y)$ customers, and the straight-line distance between customer i and customer j is L_{ij}. At time t, the pheromone concentration on the straight path between customers ij is $\theta_{ij}(t)$; assuming that the initial time $\Delta\theta_{ij}(0) = 0$, the initial iteration number $N = 0$. Then, iterate and visit customers. For each iteration and visit, the number of iterations increases by 1, and the number of ant k also increases by 1. At time t, the state probability of ants moving from customer i to customer j is $P_{ij}^{k}(t)$, which is calculated as shown in the following formula:

$$P_{ij}^{k} = \begin{cases} \dfrac{\left[\theta_{ij}(t)^{\sigma}\right] \cdot \left[\mu_{ir}(t)\right]^{\omega}}{\sum_{r \in \text{allow}_k} \left[\theta_{ir}(t)\right]^{\sigma} \cdot \left[\mu_{ir}(t)\right]^{\omega}}, & r \in \text{allow}_k, \\ 0, & r \notin \text{allow}_k. \end{cases} \tag{14}$$

The ants move to the next customer after selecting the path; calculate the total length of each ant's walking path in this section of the path and find the shortest path. Its function expression is shown in the following formula:

$$\theta_{ij}(t+1) = (1 - P)\theta_{ij}(t) + \Delta\theta_{ij}, \tag{15}$$

$$\Delta\theta_{ij} = \sum_{k-1}^{n} \Delta\theta_{ij}^{k}. \tag{16}$$

Ant colony algorithm has many advantages. For example, because it is positive feedback, the convergence speed is fast, and it is performed simultaneously in time [21]. In addition, using the principle of positive feedback, the optimization speed is fast and the optimal solution can be obtained. The disadvantage is that this algorithm is very complicated to solve the problem.

2.3.3. Particle Swarm Algorithm. In the particle swarm algorithm, the birds in the target space are the particles in the solution space, that is, the solution of the algorithm, and the solution to the problem is the food source the birds are looking for. In the optimization process of the algorithm, the particles always learn from two values: one is the individual historical optimal solution (pbest) and the other is the population historical optimal solution (gbest) [22]. Each particle improves its position and speed based on these two-dimensional values. The size and quality of the fitness value depend on the solution of the actual application problem.

In the particle swarm algorithm, the population size of particles is NP, and each particle can be regarded as a point in the solution space. The position of the first particle is w_i, and the flying speed of the particle is d_i. Therefore, the position and velocity of the particle are updated, as shown in the following formulas:

$$d_{ij}(t+1) = m \cdot d_{ij}(t) + k_1 \cdot r_1 * \left(p\text{best}_{ij}(t) - w_{ij}(t)\right) + k_2 \cdot r_2 * \left(g\text{best}_{j}(t) - w_{ij}(t)\right), \tag{17}$$

$$w_{ij}(t+1) = w_{ij}(t) + d_{ij}(t+1), \tag{18}$$

where k_1, k_2 represent the learning factor and are responsible for adjusting the step length of particle learning.

Compared with other optimization algorithms, particle swarm optimization has the same origin as them. All these algorithms need to perform a specific random search in the initial stage of optimization and obtain the next new position through its own iteration. Also, they all use fitness functions to solve individual solutions and global optimal solutions and then use evaluation criteria to obtain the optimal solutions. However, the particle swarm algorithm also has its unique features. In the analysis of the optimization formula, the particle swarm algorithm itself contains the update of speed and position. Although the global search ability is weak, the local search ability and convergence speed are very good.

3. Experiments on Solving Vehicle Routing Optimization Problem Based on Intelligent Algorithm

3.1. Model Solving Algorithm Design Based on Improved GA

3.1.1. Parameter Setting and Coding Design. After setting the parameters, first perform the genetic coding operation on the executable solution of the problem. While generating the initial population, design the fitness function that adapts to the initial population and the termination conditions of the GA according to the input conditions of the mathematical model.

Aiming at the vehicle path optimization problem that this article needs to solve, the natural number sequence is used for encoding. Assuming that x demand points are delivered by y vehicles, the length of the chromosome at this time is $x + y + 1$, and its code is shown in the following formula:

$$\left(0, n_{11}, n_{12}, \ldots, n_{1r}, 0, n_{21}, n_{22}, \ldots, n_{2t}, 0, \ldots, 0, n_{y1}, n_{y2}, \ldots, n_{yw}, 0\right), \tag{19}$$

where n_{yw} indicates that the vehicle y provides services to the customer w in the corresponding time period, 0 is the virtual distribution center, and the real distribution center is located before and after 0.

3.1.2. Initial Population and Fitness Function. The fitness function design of this article must first carry out goal transformation and constraint processing. Standardize customer satisfaction and distribution costs and then use linear weighting method to transform; the single-objective function is obtained as

$$Z = \alpha_1 \cdot z_A + \alpha_2 \cdot z_B, \tag{20}$$

where α_1, α_2, respectively, represent the weights of distribution cost and customer satisfaction ($\alpha_1 + \alpha_2 = 1$), and the weight coefficient can be adjusted according to the actual situation. This paper adopts the penalty function method to deal with the constraint conditions, and the constructed penalty function is

$$F(x, R) = f(x) - R\left[\sum_{i=1}^{m}\left[\max\left(0, k_i(x)\right)\right] + \sum_{j=1}^{l} r_j(x)\right], \tag{21}$$

where $F(x, R)$ is the penalty function. Based on the above analysis, the fitness function is

$$f = \max Z - R\sum_{i=1}^{n+m}\left(\max\sum_{j=1}^{n+m}\sum_{c=1}^{c_m}\sum_{m=n+1}^{n+m} x_{ijc}^{m} \cdot s_i - S, 0\right), \tag{22}$$

where R is an extremely positive number, and other formulas are calculation formulas for parameter constants and variable constraints.

3.1.3. Genetic Operation Design. The selection operation can be performed within an existing executable solution or within the neighborhood of an executable solution. In the GA, for the distribution problem from the distribution center to the customer point, the selection operation is mainly to find a better solution from the currently executable solution group and use it to replace the original solution group.

The crossover operation uses single-point crossover. The selected two parent individuals are randomly selected as the crossover point.

This article uses the interchange mutation method to perform mutation operations on chromosomes, select chromosomes, and randomly generate two natural numbers; if the natural number is not 0, exchange the genes corresponding to the two natural numbers to generate a new chromosome.

3.2. Experimental Simulation

3.2.1. Experimental Data and Related Parameters. To verify the effectiveness and reliability of the algorithm, this paper conducted a simulation experiment with a distribution center and randomly selected 15 nearby customer points as the distribution nodes. According to the time tolerance interval of each customer point, the time demand sensitivity is evaluated. The demand, expected time window, and tolerance time window of each customer point are shown in Table 1.

The vehicle models are uniform, with a load of 532 kg. The fixed cost for each departure is 120 yuan, and the fuel consumption per kilometer is 0.58 yuan. It is assumed that the vehicle runs at an average speed and the speed is controlled at 45 km/h.

3.2.2. Model Application Testing. Customer satisfaction analysis, according to the customer's expected time window and tolerance time window, analyzes the customer's time, demand sensitivity, satisfaction, and delivery cost.

4. Discussion on Optimization Result of Vehicle Transportation Path

4.1. Algorithm Validity. To ensure the efficiency and probability of GA improvement, in this paper, the number of configurations is set to 300, and the optimal solution results and the average running time of the improved GA, standard GA, and particle swarm algorithm are compared. The convergence diagrams of the three algorithms are as follows.

As shown in Figure 1, the three algorithms have differences in convergence speed. The improved GA in this paper began to converge in the 100th generation, the standard GA began to converge in the 150th generation, and the particle swarm algorithm began to converge in the 160th generation. This shows that the improved GA in this paper is better than the standard GA and particle swarm algorithm in convergence speed. The solution results and running time of the three algorithms are as follows.

As shown in Table 2, the running times of the three algorithms are 37.265 s, 49.338 s, and 51.646 s, the distribution costs are 498.09 yuan, 577.235 yuan, and 598.55 yuan, and the customer satisfaction is 140.45, 117.884, and 116.947, respectively. This shows that the operating time and optimization results of the GA improvement in this paper are much better than those of the standard GA and particle swarm algorithm.

4.2. Distribution Route Plan

4.2.1. Time Demand Sensitivity. According to the expected time window and tolerable time window of each customer in Table 1, the sensitivity of each customer to the time demand is calculated. The calculation results are as follows.

As shown in Table 3, each customer has his/her own time demand sensitivity and the goods must be delivered within the expected time window; otherwise, compensation will be

Table 1: Customer demand information.

Client	Demand (kg)	Expected time window	Time demand sensitivity
1	120	7:00–8:00	6:30–8:30
2	88	9:30–10:30	9:00–11:00
3	155	9:00–10:00	8:00–11:00
4	45	7:30–8:00	7:30–8:00
5	100	10:00–11:00	9:00–11:30
6	60	9:30–10:30	9:00–11:00
7	135	10:30–11:30	10:00–11:30
8	110	7:00–8:00	6:30–8:30
9	45	11:00–12:00	11:00–12:30
10	36	10:00–11:00	9:00–11:00
11	80	9:00–10:00	9:00–11:00
12	70	8:00–9:00	7:00–10:00
13	125	11:30–12:00	11:30–12:00
14	120	11:00–12:00	10:00–12:30
15	90	8:30–9:30	7:00–10:00

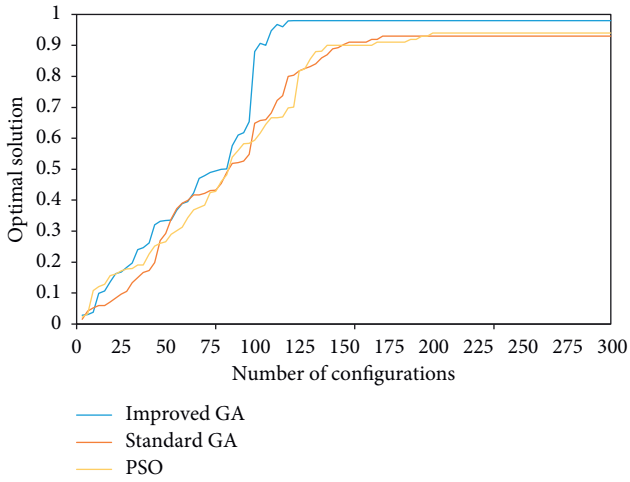


Figure 1: Convergence speed of three algorithms.

required and the distribution cost will increase. To compare the sensitivity of time demand between customers more intuitively, we draw it as a histogram.

As shown in Figure 2, we can intuitively see that customer point 4 and customer point 13 are very time sensitive, and their sensitivity is 1, while customer point 5 and customer point 12 are less sensitive to time. Therefore, it is necessary to plan the distribution plan according to the customer's time sensitivity.

4.2.2. Delivery Route. According to the geographical location and time sensitivity of the customer site, the distribution route is planned. There are three distribution vehicles with the same model, and they start from the distribution center at the same time. The distribution route plan is as follows.

As shown in Figure 3, three vehicles depart from the distribution center at the same time, go to different customer locations, and finally return to the distribution center to indicate the completion of the distribution task.

4.3. Customer Satisfaction and Distribution Costs

4.3.1. Customer Satisfaction. The satisfaction of each customer and each route is investigated, and the distribution plan and the random distribution plan under the improved GA model are used to compare the customer satisfaction of the two. The highest score of each customer satisfaction is 10 points, a total of 150 points. The specific scores are as follows.

As shown in Table 4, there is a big difference in customer satisfaction between the distribution scheme under the GA model and the random distribution scheme. Under the GA model, customer satisfaction can reach a maximum of 10 points and a minimum of 8.9 points. Under the random distribution scheme, the highest customer satisfaction score is 9.5 points and the lowest is only 8 points.

To further analyze the customer satisfaction of different vehicles under different scenarios, we compare the customer satisfaction of the customer points that each vehicle passes. Under the two scenarios, the total satisfaction of vehicle 1, vehicle 2, and vehicle 3 is compared as follows.

As shown in Figure 4, no matter which car, the satisfaction of the delivery plan under the improved GA model is higher than that of the random delivery plan. Especially for vehicle 2, the satisfaction scores of the two are 47.8 and 43 points, respectively, a difference of 4.8 points. After comparing the overall satisfaction of the vehicle, the satisfaction of each customer is compared. The comparison results are as follows.

As shown in Figure 5, although from the overall satisfaction point of view, the distribution scheme under the improved GA model is better, from the perspective of individual customers, not all individual customer satisfaction levels of the random scheme are lower than those of the algorithm scheme. Customer 6's satisfaction with the two solutions is the same (both are 9 points). Customer 8's satisfaction with the random scheme is 9.5 points, which is higher than that of the algorithm scheme by 0.5 points. Customer 15's satisfaction with the random scheme is 9.2 points, which is 1 point higher than that of the algorithm scheme.

4.3.2. Distribution Cost. In the delivery process, the cost composition of each vehicle includes a fixed cost of 120 yuan and a fuel consumption cost of 0.58 yuan per kilometer. If the vehicle arrives earlier than the expected time window, it needs to pay a penalty cost of 0.5 yuan per minute; if the vehicle arrives later than the expected time window, it needs to pay a penalty cost of 0.8 yuan per minute. Through calculation, the distribution cost of the three vehicles under the two scenarios is as follows.

As shown in Table 5, under different distribution schemes, although the fixed cost of each vehicle is the same during distribution, the fuel consumption cost and penalty cost in the actual distribution process are different, so the total distribution cost incurred is also inconsistent. The total distribution cost of the algorithm scheme is 498.09 yuan, and the total distribution cost of the random scheme is 573.13 yuan. To analyze in detail the distribution cost of three vehicles under different schemes, the distribution cost of each vehicle under the two schemes was compared.

Table 2: Solution results of three algorithms.

Algorithm	Operation hours (s)	Distribution cost	Customer satisfaction
Improved GA	37.265	498.09	140.45
Standard GA	49.338	577.25	117.884
PSO	51.646	598.55	116.947

Table 3: Customer's time demand sensitivity.

Client	1	2	3	4	5	6	7	8
Sensitivity	0.8	0.8	0.6	1	0.5	0.8	0.8	0.8
Client	9	10	11	12	13	14	15	—
Sensitivity	0.7	0.7	0.6	0.5	1	0.7	0.8	—

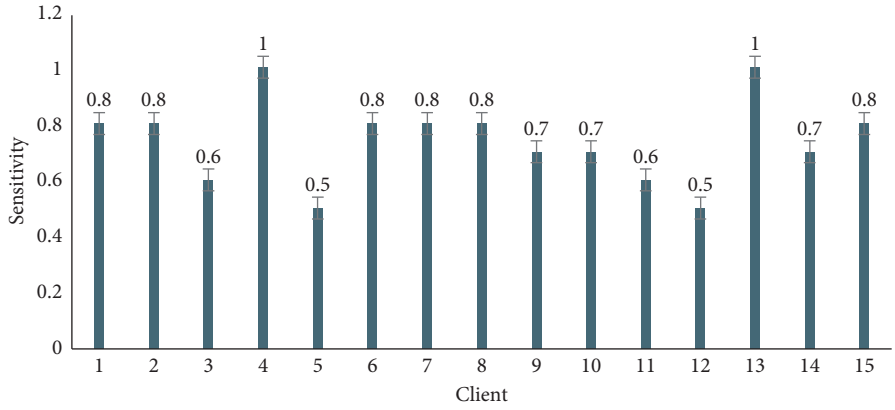


Figure 2: Customer's time demand sensitivity.

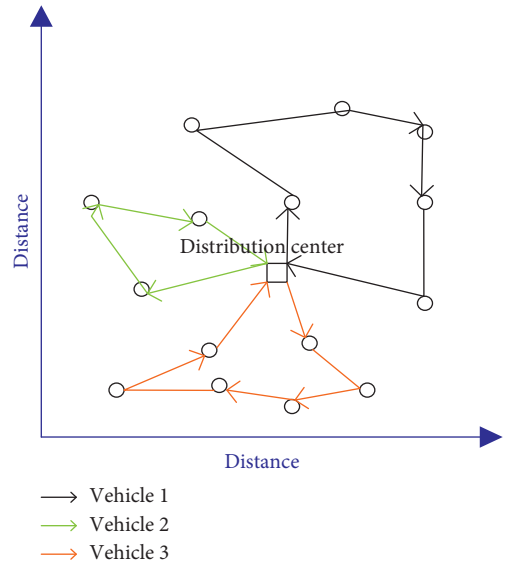


Figure 3: Vehicle distribution path.

As shown in Figure 6, whether it is fuel consumption cost or penalty cost, the distribution scheme under the improved GA model is almost lower than the random scheme. The biggest difference is the penalty cost of vehicle 3. The difference between the two schemes is 16.07 yuan. Then, we analyze the composition and proportion of the distribution cost of the improved GA, and the results are as follows.

As shown in Figure 7, in the total distribution cost of 498.09 yuan, fixed costs accounted for the largest proportion, totaling 72%, followed by the fuel consumption cost, totaling 20%, and the penalty cost is the least, accounting for only 8%. Among the three vehicles, vehicle 3 is the one with the most distribution cost, totaling 171.78 yuan, accounting for 35% of the total cost.

Table 4: Comparison of customer satisfaction.

Vehicle	Client	Algorithm solution satisfaction	Client	Random scheme satisfaction
	1	9.5	1	9
	3	10	12	9.1
1	6	9	11	8.8
	5	9.45	5	8.75
	9	9.25	9	8.5
Total		47.2	Total	44.15
	4	9.8	4	9.2
	12	9.7	3	9
2	2	10	2	8.5
	7	9.3	10	8.3
	13	9	14	8
Total		47.8	Total	43
	8	9.45	8	9.5
	15	9.1	15	9.2
3	11	9	6	9
	10	9	7	8.5
	14	8.9	13	8
Total		45.45	Total	44.2

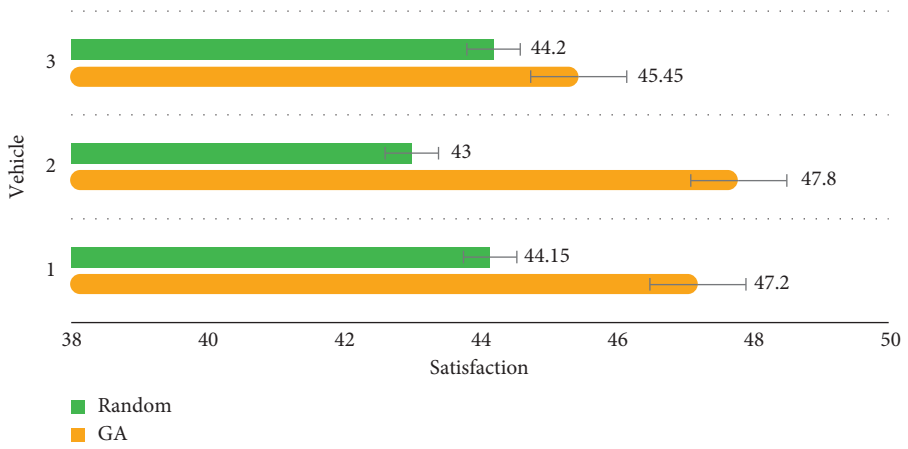


Figure 4: Comparison of customer satisfaction of different vehicles.

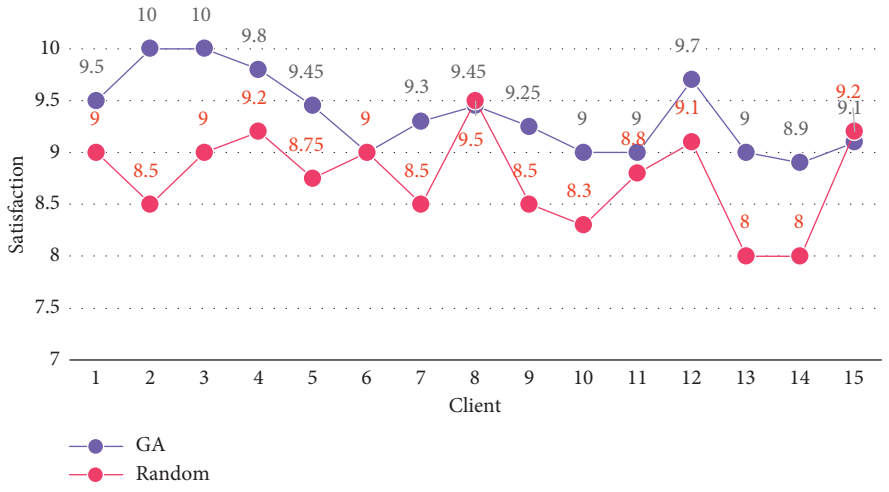


Figure 5: Comparison of satisfaction of different customers.

Table 5: Comparison of vehicle distribution costs.

Vehicle (GA)	Fixed cost	Fuel consumption cost	Penalty cost	Total
1	120	33.65	11.92	165.57
2	120	25.43	15.31	16.74
3	120	38.62	13.16	171.78
Total	360	97.7	40.39	498.09
Vehicle (R)	Fixed cost	Fuel consumption cost	Penalty cost	Total
1	120	48.74	25.24	193.98
2	120	34.01	30.69	184.7
3	120	45.22	29.23	194.45
Total	360	127.97	85.16	573.13

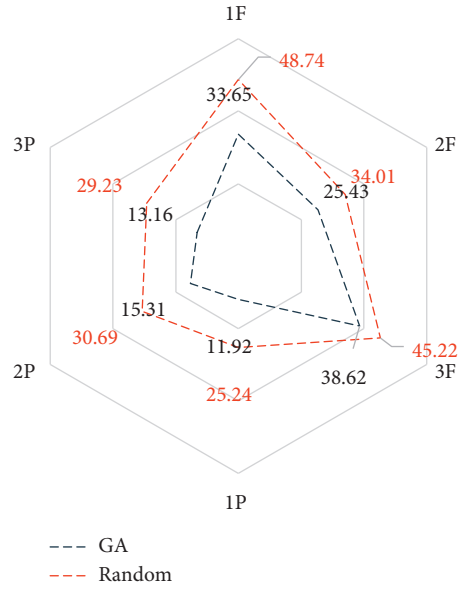


Figure 6: Distribution costs of different vehicles.

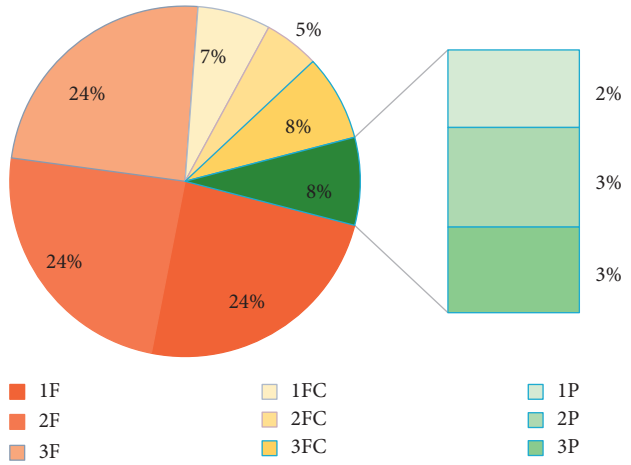


Figure 7: Distribution cost composition proportion under the algorithm scheme.

5. Conclusions

The application of IoT technology in optimizing vehicle routing can improve the reliability of the distribution process, as well as real-time tracking and monitoring of the distribution process. It can also provide key data summarized by the route optimization model, such as the state of the goods, the flatness of the route, and the time of traffic jam.

Based on the improved GA and the classic vehicle path optimization problem, the GA vehicle path optimization model established has better optimization ability and shorter

running time. By comparing distribution costs and customer satisfaction, the data show that the algorithm model established in this paper can improve customer satisfaction and save distribution costs.

Therefore, in the optimization process, it is assumed that each vehicle travels at the same speed. However, in reality, there may be unexpected phenomenon of traffic congestion, which affects the delivery time and cost to a large extent. Due to limited time and knowledge, this study did not consider this situation. Therefore, in the following research work, we will focus on the limitations of this study and make the research results closer to reality.

Data Availability

The data underlying the results presented in the study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Cao, "Research on application of the IoT technology in financial leasing of intelligent manufacturing enterprises," *International Journal of Advanced Manufacturing Technology*, vol. 107, no. 3, pp. 1061–1070, 2020.

[2] L. Hongqi, C. Xinyu, and Z. Wencong, "The vehicle flow formulation and savings-based algorithm for the rollon-rolloff vehicle routing problem," *European Journal of Operational Research*, vol. 257, no. 3, pp. 859–869, 2017.

[3] T. K. L. Hui, R. S. Sherratt, and D. D. Sanchez, "Major requirements for building smart homes in smart cities based on IoT technologies," *Future Generation Computer Systems*, vol. 76, no. nov, pp. 358–369, 2016.

[4] Z. H. Wu, "Research on the application of IoT technology to digital museum construction," *Acta Geoentica Sinica*, vol. 38, no. 2, pp. 293–298, 2017.

[5] M. R. Reddy, K. G. Srinivasa, and B. E. Reddy, "Smart vehicular system based on the internet of things," *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 45–62, 2018.

[6] B. Yao, B. Yu, P. Hu, J. Gao, and M. Zhang, "An improved particle swarm optimization for carton heterogeneous vehicle routing problem with a collection depot," *Annals of Operations Research*, vol. 242, no. 2, pp. 303–320, 2016.

[7] C. Zhou and L. Y. Ding, "Safety barrier warning system for underground construction sites using Internet-of-Things technologies," *Automation in Construction*, vol. 83, no. nov, pp. 372–389, 2017.

[8] J. Delsing, "Local cloud internet of things automation: technology and business model features of distributed internet of things automation solutions," *IEEE Industrial Electronics Magazine*, vol. 11, no. 4, pp. 8–21, 2017.

[9] D. Bo, D. Kangcheng, and C. Guang, "Carbon emission management system of port logistics based on IoT technology," *Agro Food Industry Hi-Tech*, vol. 28, no. 1, pp. 1094–1098, 2017.

[10] Y. Zhang and J. Wen, "The IoT electric business model: using blockchain technology for the internet of things," *Peer-to-Peer Networking and Applications*, vol. 10, no. 4, pp. 983–994, 2017.

[11] E. Zachariadis, E. Emmanouil, C. D. Tarantilis, and C. T. Kiranoudis, "The pallet-packing vehicle routing problem," *Transportation Science*, vol. 46, no. 3, pp. 341–358, 2012.

[12] S. Wang, Z. Lu, L. Wei, G. Ji, and J. Yang, "Fitness-scaling adaptive genetic algorithm with local search for solving the Multiple Depot Vehicle Routing Problem," *Simulation*, vol. 92, no. 7, pp. 601–616, 2016.

[13] S. Wan, X. Li, and Y. Xue, "Efficient computation offloading for Internet of Vehicles in edge computing-assisted 5G networks," *The Journal of Supercomputing*, vol. 76, no. 4, 2019.

[14] A. Rautela, S. K. Sharma, and P. Bhardwaj, "Distribution planning using capacitated clustering and vehicle routing problem," *Journal of advances in management research*, vol. 16, no. 5, pp. 781–795, 2019.

[15] W. Li, Y. Wu, and P. N. R. Kumar, "Multi-trip vehicle routing problem with order release time," *Engineering Optimization*, vol. 52, no. 10, pp. 1–16, 2019.

[16] D. W. Gong, J. Sun, and Z. Miao, "A set-based GA for interval many-objective optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 99, pp. 47–60, 2018.

[17] T. Wang, Z. Zheng, and M. Elhoseny, "Equivalent mechanism: releasing location data with errors through differential privacy," *Future Generation Computer Systems*, vol. 98, no. Sep, pp. 600–608, 2019.

[18] H. Aziza and S. Krichen, "Bi-objective decision support system for task-scheduling based on genetic algorithm in cloud computing," *Computing*, vol. 100, no. 2, pp. 65–91, 2018.

[19] Y. Feng, M. Zhou, G. Tian et al., "Target disassembly sequencing and scheme evaluation for CNC machine tools using improved multiobjective ant colony algorithm and fuzzy integral," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 12, pp. 2438–2451, 2019.

[20] D. Shu, Z. Huang, J. Li, and X. Zou, "Application of multi-agent particle swarm algorithm in distribution network reconfiguration," *Chinese Journal of Electronics*, vol. 25, no. 6, pp. 1179–1185, 2016.

[21] Y. Cheng and J. Wu, "Particle swarm algorithm-based damage-mitigating control law analysis and synthesis for liquid-propellant rocket engine," *Proceedings of the Institution of Mechanical Engineers-Part G: Journal of Aerospace Engineering*, vol. 233, no. 10, pp. 3810–3818, 2019.

[22] S. Almahdi and S. Y. Yang, "A constrained portfolio trading system using particle swarm algorithm and recurrent reinforcement learning," *Expert Systems with Applications*, vol. 130, no. SEP, pp. 145–156, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 4046083, 11 pages
https://doi.org/10.1155/2021/4046083

Research Article

Missing Information Reconstruction of Land Surface Temperature Data Based on LPRN

Chen Xue,[1] **Tao Wu**,[1] **Xiaomeng Huang**,[1,2] **and Amir Homayoon Ashrafzadeh**[3]

[1]*School of Computer Science, Chengdu University of Information Technology, Chengdu, China*
[2]*Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China*
[3]*RMIT University, Melbourne, Australia*

Correspondence should be addressed to Tao Wu; wut@cuit.edu.cn

Received 19 August 2021; Accepted 6 September 2021; Published 21 September 2021

Academic Editor: Xianyong Li

Temperature is the main driving force of most ecological processes on Earth, with temperature data often used as a key environmental indicator to guide various applications and research fields. However, collected temperature data are limited by the hardware conditions of the sensors and atmospheric conditions such as clouds, resulting in temperature data that are often incomplete. This affects the accuracy of results using the data. Machine learning methods have been applied to the task of completing missing data, with mixed results. We propose a new data reconstruction framework to improve this performance. Using the MODIS LST map over a span of 9 years (2000–2008), we reconstruct the land surface temperature (LST) data. The experimental results show that, compared with the traditional reconstruction method of LST data, the proportion of effective pixels of the LST data reconstructed by the new framework is increased by 3%–7%, and the optimization effect of the method is close to 20%. The experiment also discussed the influence of different altitudes on the data reconstruction effect and the influence of different loss functions on the experimental results.

1. Introduction

As a key element of biological survival, temperature is an important subject in the field of climate research. Researchers can determine the water-heat balance between the Earth's surface and the atmosphere based on surface temperature data. These data can also be used as a basis for understanding various terrestrial activities, determining fire and seismic zones, exploring geothermal resources, and studying urban heat island effects [1, 2]. Temperature also directly affects evapotranspiration and soil content and is key to evaluating terrestrial water volume, vegetation, and soil biochemical characteristics; atmospheric precipitation; and regional CO_2 content [3–9]. Temperature data are a common input variable into models of atmospheric, ecological, hydrological, and biogeochemical processes. The accuracy of the data directly affects the output accuracy and overall model accuracy [10, 11]. Therefore, it is important to obtain high-precision, continuous surface temperature data.

The collection of surface temperature data was initially carried out on mobile phones through discrete observation stations, obtaining high-precision, all-weather data through a large number of ground observation stations. However, the ground stations do not completely cover the Earth's surface, leading to a loss of data in many areas. Technological advances have led to Earth observation satellites as the mainstream data collection method, with remote sensing technology the only surface temperature observation method able to guarantee thorough coverage at all times in all areas on a global scale [12]. The collection of surface temperature using remote sensing technology is now an effective method for obtaining surface temperatures at a large scale. The theory and method of land surface temperature inversion from thermal infrared remote sensors are now relatively mature and produce high-precision clear-sky surface temperatures [13, 14]. However, thermal infrared remote sensing cannot penetrate clouds, leaving areas under cloud cover inaccessible and affecting data collection.

Solving the problem of missing temperature data due to cloud cover is currently a major topic in the field of climate research.

So far, researchers have applied many mature data reconstruction methods to solve the problem of temperature information loss. These methods fall into three broad categories [15]: space-based methods, spectrum-based methods, and time-based methods. Details of these methods are provided in the discussion in Section 2.2. In the research on the reconstruction of land surface temperature data, Shuo Xu et al. reconstructed the land surface temperature in the Tibetan Plateau (TP) and the Heihe River Basin (HRB) area based on the Bayesian maximum entropy (BME) method and verified the effectiveness of the method by using the soil temperature measured in the field [16]. Zhang et al. merged TIR and MW observations from a perspective of decomposition of LST in temporal dimension and overcame shortcomings of single-source remote sensing [17]. Martins et al. proposed an all-weather LST product based on visible light and infrared observations, which combines clear-sky LST retrieved from the Spinning Enhanced Visible and Infrared Imager on Meteosat Second Generation (MSG/SEVIRI) infrared (IR) measurements with LST estimated with a land surface energy balance (EB) model to fill gaps caused by clouds [18]. Although these different methods can acquire satisfactory recovery results, most of them are employed independently, and they can only be applied to a single specific reconstruction task in limited conditions [19]. On the one hand, most traditional data reconstruction methods such as Bayesian and KNN interpolation are based on linear regression, which requires a linear relationship between the data. However, the actual data missing scene is more complicated, and the data are often nonlinearly related. On the other hand, the ability of traditional data reconstruction methods to process data is limited. When the data reach a certain amount, the time to process the data will increase rapidly, and the data reconstruction experiment will be greatly affected. Therefore, it is necessary to find a better data reconstruction method that can integrate diversified information to overcome the limitations and deficiencies of traditional methods.

In recent years, benefiting from the improvement of computer computing power, LeCun et al. made a huge breakthrough in the field of image recognition using machine learning methods [20–23]. Considering that the satellite remote sensing field also needs to complete the recognition and classification of remote sensing satellite images and other similar tasks, related scholars have tried to introduce machine learning methods into the field of Earth sciences and have achieved certain results. Chen et al. used the Apriori method to find potential correlations in geological data [24]. Xie and Li used deep learning methods to denoise hyperspectral images [25]. Wei et al. improved the accuracy of sharpened images through the residual network [26]. Shah et al. proved the correlation between LST anomalies and earthquakes (EQs) in Pakistan with the ANN method based on MODIS LST data [27]. However, in terms of the reconstruction task of LST data based on machine learning methods, it is still difficult to find related research.

Therefore, we propose the LST palindrome reconstruction network (LPRN) method, which uses a deep convolutional neural network to reconstruct remote sensing images contaminated by dense clouds in MODIS LST data. At the same time, in order to provide high-quality training data and label data for the deep learning framework, the LPRN method also incorporates a variety of data preprocessing schemes, such as histogram equalization, inverse distance weighted interpolation, and so on. Compared with traditional data reconstruction methods, this method can effectively improve the number of "good quality" pixels (also represents effective pixels, explained in Section 2.1) in cloud-contaminated images while also greatly improving computing efficiency.

The rest of this article is organized as follows. In Section 2, we introduce the available datasets for reconstruction, current mainstream methods for reconstructing remote sensing data, and the network structure and specific details of our LPRN data reconstruction model. Section 3 presents our experimental results from using the model to reconstruct the surface temperature dataset. Finally, Section 4 presents our summary, concluding remarks, and directions for future research.

2. Materials and Methods

2.1. Datasets. The dataset used in the experiment came from the Terra-MODIS and the MODIS-Aqua datasets (https://modis.gsfc.nasa.gov/data/dataprod/mod11.php), and the MODIS data product IDs used are MOD11A1, MYD11A1, MOD11A2, and MYD11A2. The Terra-MODIS dataset was from March 2000, and MODIS-Aqua from August 2002, both with daily resolution. The time difference between the Aqua satellite data and the Terra satellite data in the same area can be as short as a few hours, so there is great similarity between the two data. And Terra satellite data can be added to the dataset as supplementary information to the Aqua satellite data. We selected and processed nine years of MODIS Daily data and 8-Day data from 2000 to 2008 as experimental data. MODIS Land Surface Temperature and Emissivity products map land surface temperatures and emissivity values ideally under clear-sky conditions [28]. The underlying algorithms use other MODIS data and further auxiliary maps for input, including geolocation, radiance, cloud masking, atmospheric temperature, water vapor, snow, and land cover [29]. Temperatures are provided in Kelvin. The MODIS LST algorithm is aimed at reaching a better accuracy than 1 Kelvin (±0.7 K stddev.) for areas with known emissivities in the range −10°C to 50°C [13, 29]. LST is observed by the two MODIS sensors four times per day (01 : 30, 10 : 30, 13 : 30, and 22 : 30, local solar time) originally at 1000 m pixel resolution. Clouds and other atmospheric disturbances, which may obscure parts of or even the entire satellite image, constitute a significant obstacle for continuous LST monitoring; the low-quality pixels of each LST map are marked in an accompanying quality assessment (QA) layer. For the LST maps, we used the open-source software GRASS to remap the MODIS LST data, filter out invalid pixels, and reject the pixels with the following labels

indicated in the QA bitmap: "other error," "missing pixel," "poor quality," "average emissivity error >0.04," "LST error 2 K–3 K," and "LST error >3 K," and the rest are "good quality" pixels. Based on the tags in the bitmap, it is possible to accurately filter the elements in the original LST image and improve the efficiency of LST data preprocessing. It should be noted that the data will only be stored in the database after being processed by the GRASS software. The dataset in the database is still in an unclassified state and there are abnormal values, and data cleaning and data classification need to be completed through the data preprocessing process.

2.2. Existing Remote Sensing Data Reconstruction Methods

Space-Based Methods. Space-based methods were originally used for image reconstruction in the field of computer vision. Among them, the most commonly used method is linear interpolation, which is weighted based on the surrounding information around missing data to estimate the missing information. This method often relies on the correlation and connection between missing information and its surrounding information. However, when the missing information changes greatly compared to the surrounding information, such as boundary values, it often results in inaccurate interpolation results. This method is reasonable, but it is limited by the amount of missing data and regional characteristics.

Spectrum-Based Methods. In the case of multispectral or hyperspectral images, specific spatial correlations between them have led some scholars to propose an innovative scheme for reconstructing missing information using these correlations. Rakwatin et al. used a polynomial linear fitting (LF) method to fit the missing data in the Aqua MODIS data [30]. Chen et al. used histogram matching and local least squares fitting methods to reconstruct Aqua MODIS data [31]. However, this method provides no help for satellite data with cloud cover.

Temporal Methods. Satellite remote sensing data are obtained through continuous satellite scanning of a defined area, with the collected data including the time of collected data. Some methods use these time elements to restore and reconstruct missing information. For example, Chen et al. proposed the use of a space-time weighted regression model (STWR) to obtain continuous cloud-free LST images [31]. Scaramuzza and Barsi proposed a local linear histogram matching (LLHM) method, but in most cases, it did not achieve good results [32]. However, it is undeniable that time differences are quite influential and useful when reconstructing missing information.

2.3. Reference Deep Learning Model

Convolutional Neural Network. CNN [23] is a multilayer neural network originally used for image recognition and classification. CNNs are generally composed of four parts: convolutional layer, nonlinear layer, pooling (also called downsampling) layer, and fully connected layer. The convolution operation primarily extracts features in the convolutional layer. The pooling layer reduces the number of calculations in the entire network and prevents overfitting. At the end of the network is the fully connected layer. The size of the convolution kernel is very important, as it determines the size of the neuron's receptive field. If the kernel is too small, it is difficult to extract effective regional features. If the kernel is too large, the network complexity increases and may exceed the representation ability of the kernel. By convolving the input feature block x with the kernel and then passing the activation function $f(x)$ plus a bias term b, the output y is obtained. The entire convolution operation is given in equation (1). The function of the pooling layer is to reduce the amount of output data, obtaining lower dimensional features, reducing the number of calculations, and preventing feature overfitting. More specifically, neurons in the network are temporarily discarded according to a certain probability to prevent overfitting and improve the generalization ability of the model. Then, added regularization can make the parameters sparser, which makes one part of the optimized parameters become 0 and other part a nonzero real value. The nonzero real value plays a role in selecting important parameters or feature dimensions and at the same time plays a role in removing noise, which can further improve the performance of the model.

$$y = f(w x + b). \tag{1}$$

Fully Convolutional Network. FCNs [33] were first used for semantic segmentation, which differs from ordinary classification tasks that output a specific category only. FCNs require the output and the input to be the same. The structure of FCNs differs from CNNs as well. On the basis of CNNs, FCNs adjusted the last layer of the network and added a "deconvolution layer." The fully connected layer of a CNN loses some of the position information in the input data. FCNs work to preserve this information by replacing the fully connected layer with a similar "deconvolutional layer." After convolution and pooling, the feature matrix is restored to the original data dimensions after the "deconvolution layer" and upsampling method. Upsampling is the inverse process of the pooling layer introduced earlier which restores data features extracted through convolution and pooling to the appropriate dimensions to protect the original structure of the data at least in part. The entire network performs a total of five pooling operations, with feature fusion performed in the last step of the network. Feature fusion requires upsampling the feature map with the smallest receptive field obtained after convolution and pooling and then fusing the result with the feature map of the previous layer. The fused feature map is again upsampled with the

feature map of the previous layer, and feature fusion is performed until the feature map is restored to the original number of data dimensions.

U-Net. U-Net [34] is a variant of the FCN network, with a changed encoder-decoder structure. There is no fully connected layer in the network, and the structure is divided into two parts: the encoder path and the decoder path. The structure of the encoder path is similar to the traditional CNN structure, and the decoder path is the encoder path's inverse process. The whole structure resembles the letter U, incorporated into the name. The encoder path mainly performs feature extraction and dimensionality reduction. The decoder path is composed of a convolutional layer, an activation layer, and an upsampling layer, which fuse features and restore the data dimensionality. The biggest difference between U-Net and FCN is in the deep and shallow feature information fusion of the data. FCN fuses feature information by adding the values at the corresponding positions, while U-Net fuses using same dimensional splicing. The latter works by placing feature information at different levels in different channels, splicing it together, and then convolving to fuse features. The FCN network fuses data features through addition, which will obscure the details of the data features, while U-Net's splicing structure can retain more position information and reduce the loss of data details.

2.4. Proposed LST Palindrome Reconstruction Network (LPRN). Inspired by the encoder-decoder framework in U-Net, we try to improve the original network structure of U-Net and propose the LPRN framework. This framework applies deep convolutional neural networks to LST data reconstruction research, and the overall architecture is shown in Figure 1. The whole framework includes three processes, namely, the reprojection of MODIS LST (explained in section *Datasets*), the preprocessing of the dataset used by the model, and the structure design of the LPRN model.

2.4.1. Data Preprocessing. The LST data after reprojection can be divided into two categories according to whether they are occluded by clouds: cloud-containing images and cloud-free images. After the cloud-containing image is filtered, the "good quality" pixels will be greatly reduced and cannot be directly used as the training dataset of the deep learning network (when there are too many missing values in the training dataset, the deep learning method cannot capture the potential correlation between the data, nor can it reconstruct the missing real data). Therefore, it is necessary to use data processing methods to interpolate the missing data to a certain extent. In the process of data preprocessing, we first use the histogram equalization (HE) method to determine whether the image is occluded by clouds, then divide the dataset, and finally use the inverse distance weighting (IDW) method (see Figure 2) to interpolate part of the missing data.

The data preprocessing process includes the following steps:

Step 1: according to the experimental needs, download the LST dataset within a specified time scale in batches, classify and store the data according to the satellite, date, and other factors, and establish the corresponding association between the data.

Step 2: divide the cloud-containing image according to the spectral characteristics—take out the image data from the database and convert it into the form of gray frequency histogram; analyze the gray frequency histogram; if two peak frequencies appear and increase with the gray value and the gray value between the peaks has a buffer zone, the gray value of the area corresponding to the second peak is set as an abnormal value and the image is divided into cloud-containing image; if the above requirements are not met, the image is classified as pending image data; the above process is repeated until all the images in the dataset are completely divided.

Step 3: analyze the spectral characteristics of the cloud-containing image through the gray frequency histogram. According to the analysis of the existing data, it is found that the frequency histogram of the nonthin cloud-containing image will show two obvious gray scale peaks. The peak frequency with the higher gray scale will be significantly higher than the one with the lower gray scale and there will be a certain buffer in the frequency between two gray values. This feature is an important basis for determining whether an image is a cloud-containing image. If the image is a cloud-free image, the gray frequency histogram of the image will show an approximate standard normal distribution. According to the above feature, the remote sensing data of LST are preliminarily classified.

Step 4: perform histogram equalization processing on the filtered pending image data and filter the cloud-containing image for the second time according to the processing result—take out the image data in sequence from the pending image dataset, perform the histogram equalization processing on the image data, and calculate the gray average value and second-order moment of the image before and after processing; classify the image data according to equations (2) and (3); if two inequalities are satisfied at the same time, the image will be classified as a cloud-containing image from the pending image dataset; otherwise, the image will be classified as a cloud-free image (where Mean_b represents the average value of the candidate image after equalization processing, Mean_f represents the average value of the candidate image before equalization processing, ASM_b represents the second moment of the candidate image after equalization processing, and ASM_f indicates that the candidate image is equalizing second moment before transformation).

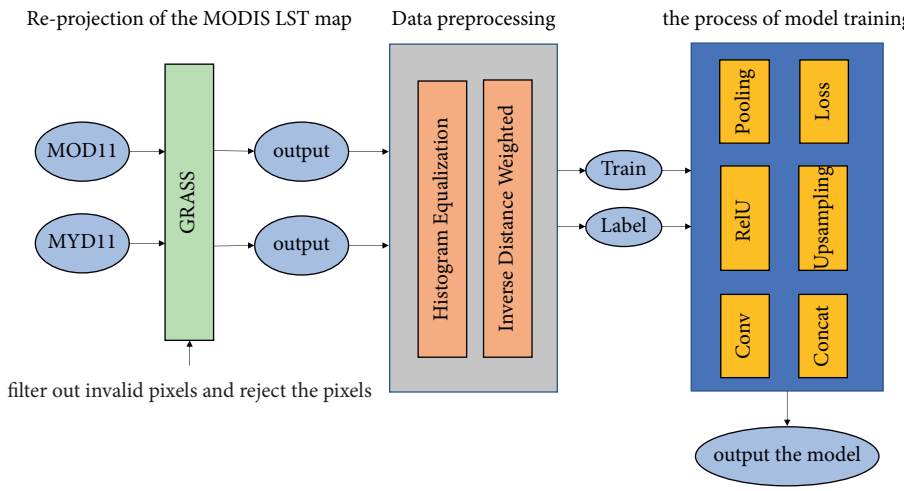


Figure 1: Flowchart of the LPRN framework for the missing information reconstruction of LST.

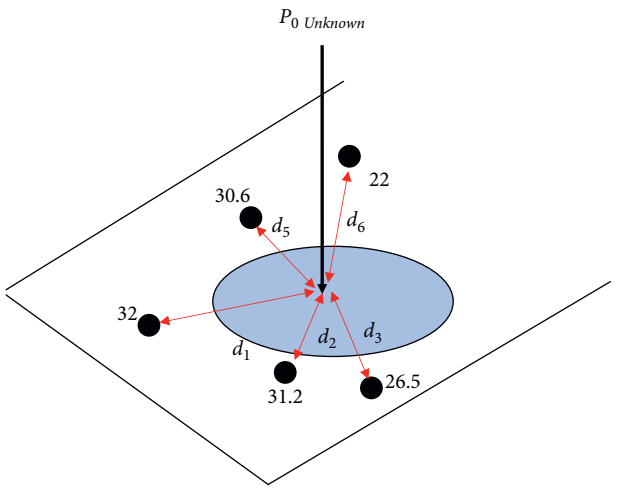


Figure 2: The inverse distance weighting.

$$2.3 < \frac{\text{Mean}_b}{\text{Mean}_f} < 3$$
$$-0.03 \leq \text{ASM}_f - \text{ASM}_b \leq 0.03. \quad (2)$$

Step 5: record the cloud occlusion range of the cloud-containing image, divide the cloud occlusion area, and count the cloud occlusion ratio—for the data classified as cloud-containing images in step 2, the cloud occlusion range is divided according to the highest gray value peak of the cloud-containing image data in the frequency histogram, and the pixels in the abnormal gray scale peak interval are all cloud occlusion areas, and count the number of cloud pixels; for the data classified as cloud-containing images in step 4, the cloud occlusion range is divided according to the average gray value of the buffer area in the distribution of the frequency histogram. If there is a gray peak value that is much larger than the average value, the area is divided into a cloud occlusion area, and count the amount of cloud pixels.

Step 6: filter data according to the cloud occlusion ratio of each sample in the cloud-containing images—set the threshold according to the proportion of cloud-occluded pixels calculated in step 5 and determine whether to accept the land surface temperature data according to the threshold (the threshold needs to be flexibly determined according to the actual amount of data, and if the threshold is set too high, too much data will be deleted).

Step 7: take out the cloud-containing image in turn and use the inverse distance weighting method to reconstruct the pixels in the cloud occlusion area according to equations (4), (5), and (3); if the range of the missing pixels is large, the pixel value of the missing area of the picture is reconstructed according to the ratio of the gray values between the pixels in the image at different moments in the same area (where d_i represents the distance from the target pixel to the surrounding pixels, λ_i represents the calculated weight based on the reciprocal of the distance, and $\widehat{P}(x, y)$ represents the sum of the products of the surrounding pixels and the weight).

$$d_i = \sqrt[2]{(x - x_i)^2 - (y - y_i)^2},$$
$$\lambda_i = \frac{1/d_i}{\left(\sum_{i=1}^{n} 1/d_i\right)}, \quad (3)$$
$$\widehat{P}(x, y) = \sum_{i=1}^{n} \lambda_i P(x_i, y_i).$$

The main purpose of the above process is to eliminate outliers and obtain the standard dataset for the LPRN model, and the brief process is shown in Figure 3.

The above method can remove the data missing problem caused by cloud occlusion to a certain extent and reduce the numerical error caused by cloud occlusion. However, in most cases, because the range of missing values is relatively large, the effect of interpolation methods based on space and time is limited.

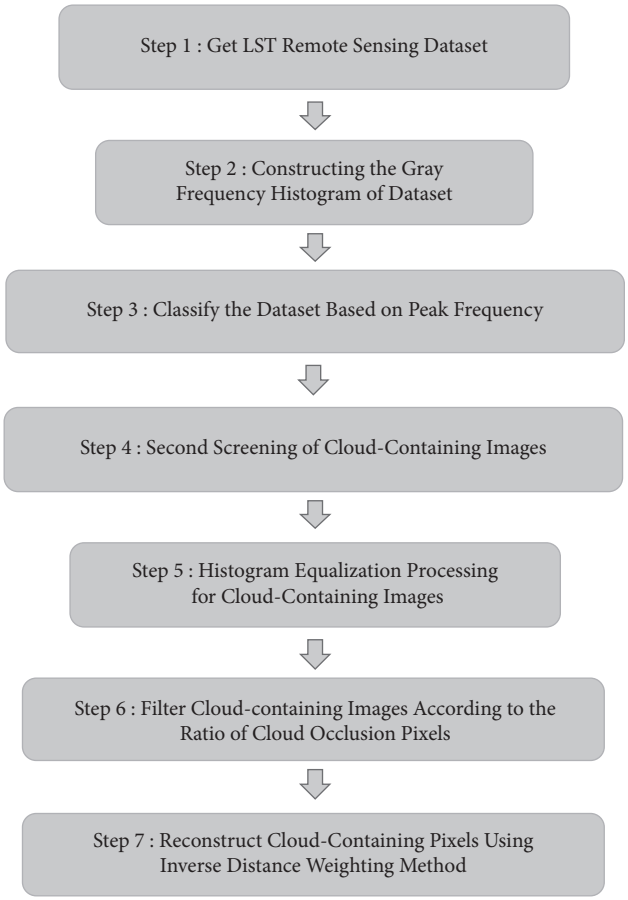


Figure 3: The flowchart of data preprocessing.

2.4.2. LPRN Framework. With reference to the structure of U-Net, the LPRN model proposed in the paper is based on the encoder-decoder architecture to reconstruct the missing data of LST. The LPRN model consists of 14 convolutional layers, 3 pooling layers, and 3 upsampling layers, and since the data characteristic specifications obtained by each layer are different, each layer has constraints on the specifications of the input data and the output data. Considering the overall calculation amount of the model, the characteristics that the model needs to obtain, and the LST data at different times and dates, the model uses convolution kernels of different sizes to obtain the characteristics of the LST data as comprehensively as possible. The overall architecture of the LPRN framework is shown in Figure 4. The label data in the proposed model are the original 8-Day LST image, and the training data in the proposed model are the preprocessed Daily LST image.

As shown in the framework, the slice size of the input Daily LST data sample is 200 ∗ 248. In the framework, five channels of different sizes are set, which are 1, 8, 16, 32, and 64. Three convolution kernels of different sizes are set to realize feature extraction of different scales, which are [3, 3], [2, 2], and [1, 1]. In order to avoid the loss of too much detailed information in the data, the downsampling layer only uses a convolution kernel with a size of [2, 2] to compress the data volume features. The framework adopts a decoder-encoder structure to effectively reduce the loss of original data information due to pooling and convolution. This structure retains the intermediate information of the data during the decoding process and then uses convolution and data splicing methods to fuse the original information and the extracted features during the encoding process, which can reconstruct the missing data according to the characteristic correlation information in the original data.

In the deep learning network, the loss function is used to estimate the degree of inconsistency between the predicted value $f(x)$ of the model and the true value Y. The smaller the loss function, the better the stability of the model. Relying on the guidance of the loss function, the deep learning model can realize the learning process.

In the experiment, we tried a variety of different loss functions. Among them, the original loss function of the U-Net network is the cross entropy function, and the commonly used loss functions of deep learning networks include mean square error (MSE) and root mean square error (RMSE). However, the experimental results show that aforementioned loss function has very limited guiding effect on LPRN, so we tried the relatively unpopular normalized cross correlation (NCC) loss function and achieved relatively good results (the experimental results of different loss functions are provided in the discussion in Section 3).

RMSE: RMSE is often used as a standard for measuring the prediction results of machine learning models, which is expressed as follows:

$$\mathrm{RMSE}(X,h) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(h(x_i) - y_i\right)^2}, \tag{4}$$

where X represents the training sample, m represents the number of training samples, $h(x)$ represents the predicted value of the training sample, and y represents the true value of the training sample, which is also called the label value.

NCC: NCC is primarily used with image matching, and it is the process of finding the subimage having greatest similarity to a real time image for the purpose of identifying a target image, which is expressed as follows:

$$\rho(x,y) = \frac{\sigma\left(S_{x,y}, g\right)}{\sqrt{D_{x,y}D}}, \tag{5}$$

where $S_{x,y}$ and g, respectively, represent the corresponding subblocks of the two compared images, $\rho(x,y)$ represents the correlation coefficient, which is used to determine whether the two subblocks are related, $\sigma(S_{x,y}, g)$ represents the covariance of $S_{x,y}$ and g, $D_{x,y}$ represents the variance of datum $S_{x,y}$, and D represents the variance of datum g.

The inequality $|\rho(x,y)| \le 1$ is used to measure the correlation between two subblocks, and the value of similarity is within [−1, 1]. The correlation coefficient characterizes the description of the degree of approximation between the two

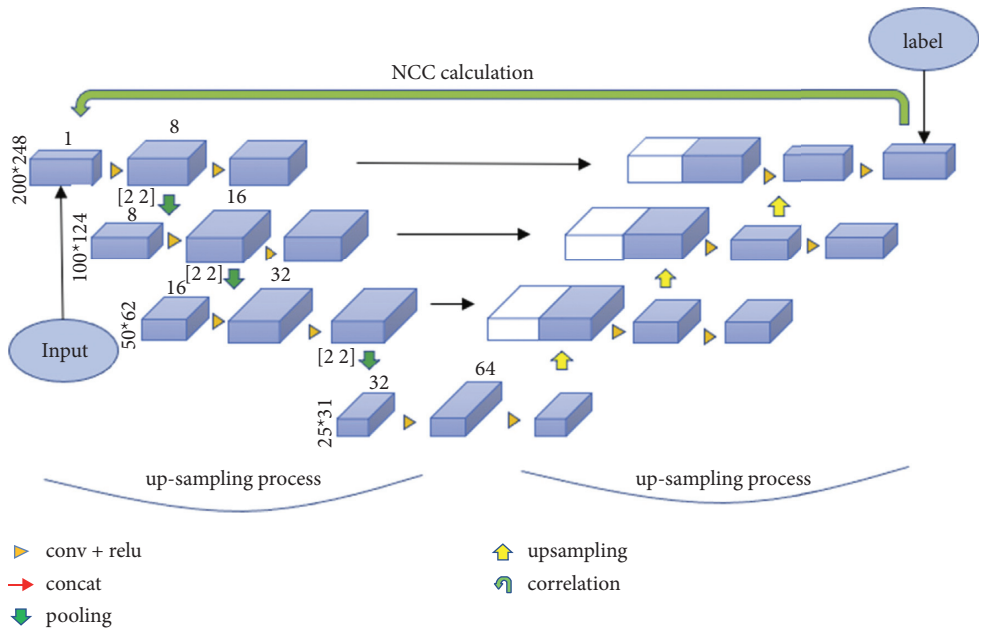


Figure 4: Architecture of the proposed LPRN framework.

data. Generally speaking, the closer it is to 1, the closer the linear relationship between the two data is.

3. Results and Discussion

Uncertainty in the processed MODIS LST data as proposed in this paper is of two types: (1) intrinsic uncertainty in the downloaded MODIS data and (2) error introduced by processing. The former includes error in the radiometric and geometric precision of the MODIS instrument and uncertainties in the known emissivity values of land surfaces, cirrus, or other atmospheric phenomena. In the experiment, the reprojection of MODIS LST and the preprocessing of the dataset are used to minimize the impact of related errors. The latter is mainly the potential error caused by the calculation of the average value in the preprocessing process and the data reconstruction process. This error may be reduced in the future by using additional data sources.

In order to comprehensively evaluate the impact of uncertainty on the accuracy of LPRN reconstruction data, this study carried out two different LST reconstruction data accuracy evaluation experiments.

In the first experiment, we mainly examine the change of a single LST value before and after reconstruction. The experiment randomly selected 2000 LST original samples from the cloud-free image, so as to avoid selecting too many similar regions. Then, we randomly select 5% of the "good quality" pixels from each original sample and modify them as missing data and record the location of the modified pixels. Finally, we put the "modified" LST original sample into the LPRN model for reconstruction and find the reconstructed LST data according to the position of the "modified" pixel and compare it with the actual value of the LST in the original sample (see Table 1) (because the images are randomly selected, the distribution of the number of images in each year is not the same).

Analyzing the results of the first experiment and comparing the reconstructed data from 2000 to 2002 with the original image, the average temperature error between the corresponding pixels is larger, and the average error is above 1 K. Comparing the reconstructed data from 2003 to 2008 with the original image, the average temperature error between the corresponding pixels is small, the average error is below 0.4 K, and the smallest error is close to 0.1 K. Analysis of the reasons for this phenomenon is mainly due to the different reconstruction effects caused by the amount of training data. The data from 2003 to 2008 are compared with the data from 2000 to 2002. The latter lacks MODIS-Aqua data, which leads to the reduction of data associations that can be referred to by the model, so the reconstruction effect of the model also decreases accordingly.

The first experiment mainly evaluates the influence of the LPRN model on the reconstruction of the original data, and the result proves that the model will not cause a large change in the normal data. However, for the newly created information in the missing area, since they are data that do not exist in the original dataset, their accuracy cannot be verified, so we introduced a data correlation coefficient $\rho(T_{re}, T_{tr})$ in the second experiment, as shown in equation (6). In the second experiment, based on the correlation between the reconstructed data and the label data, we further verify the accuracy of the reconstructed data of the LPRN model (experiment 2 used all the extracted cloud-free data into the LPRN model for experimentation, and because there are few cloud-free data available from 2000 to 2002, they were not used in the experiment).

Table 1: Average error statistics of the single pixel.

Year	Error (K)
2000	1.3
2001	1.4
2002	1.0
2003	0.4
2004	0.1
2005	0.3
2006	0.4
2007	0.2
2008	0.3
Mean	0.6

$$\rho(T_{re}, T_{tr}) = \frac{Cov(T_{re}, T_{tr})}{\sqrt{D(T_{re})}\sqrt{D(T_{tr})}}, \quad (6)$$

where T_{re} and T_{tr}, respectively, indicate reconstructed data and label data and $\mathrm{Cov}(\cdot)$ and $D(\cdot)$ indicate covariance and variance, respectively.

The second experiment first finds out the corresponding data in the label dataset according to the date of the cloud-free data and then puts all the cloud-free data into the trained LPRN framework for reconstruction and obtains the reconstructed cloud-free dataset. Finally, the correlation and error between the original cloud-free dataset, the reconstructed cloud-free dataset, and the label dataset are calculated to verify the accuracy of the reconstructed data (see Table 2).

The focus of the introduction of the correlation coefficient is to test the accuracy of the reconstruction results of the LPRN model. Analyzing the data results in Table 2, we first analyze the correlation between the original data and the label data. The maximum correlation coefficient appeared in 2004 and was 0.90, and the error in that year was also the smallest, with an average error of 2.92 K. The correlation coefficient in 2006 was the smallest, 0.87, and the error in that year was also the largest (5.07 K). The average value of the correlation coefficient between the original data and the label data is 0.883, and the average value of the error is 4.265. Then, we analyze the correlation coefficient and error between the LST data reconstructed by the LPRN method and the label data. The maximum value of the correlation coefficient is 0.93, the minimum value is 0.90, the minimum error is 1.21 K, and the maximum value is 2.62 K. After the reconstruction of the LST data, the average value of the correlation coefficient is improved by nearly 0.03, and the error is reduced by 2.4 K on average.

The results of experiment two show that the correlation coefficient and the error are linearly related, and the smaller the error, the greater the correlation. Analyzing the error between the original data before and after reconstruction and the label data, it can be determined that the reconstruction effect of the LPRN model is obvious. After comprehensive analysis of the results of the first experiment and the second experiment, the LPRN model can significantly reduce the error between the original data and the label data. But when the correlation coefficient of the two data reaches a certain value, the model reconstruction effect will be limited. Therefore, more diversified label data can be considered in future research, and the reconstruction effect of the LPRN model can be improved by reducing the correlation between the data. At the same time, it should be noted that, compared with the first experiment, as the number of verification data increases, the error will also increase to a certain extent. This may be due to the small amount of data used in the first experiment, resulting in greater randomness in the results.

Table 2: Evaluations of the reconstructed data.

Year	$\rho(T_{ori}, T_{tr})$	Error (T_{ori}, T_{tr})	$\rho(T_{re}, T_{tr})$	Error (T_{re}, T_{tr})
2003	0.88	4.90 K	0.92	1.68 K
2004	0.90	2.92 K	0.93	1.21 K
2005	0.88	4.31 K	0.91	2.15 K
2006	0.87	5.07 K	0.92	1.74 K
2007	0.89	3.83 K	0.90	2.62 K
2008	0.88	4.56 K	0.93	1.43 K
Mean	0.883	4.265 K	0.918	1.805 K

In addition, in order to test the effect of the LPRN model on the reconstruction of missing data, the experiment selects the method proposed in [35] for comparison. Reference [35] mainly uses traditional data reconstruction methods such as nearest neighbor algorithm and histogram equalization. Considering that the experimental results are all reconstructed data and lack the reference of real data, the experiment mainly compares the number of "good quality" pixels in the image reconstructed by the two methods. The experiment selects the same area as in research [35] (46 : 45 : 00 N-45 : 36 : 50 N; 10 : 06 : 33E-12 : 46 : 30E) and data size (original size after data download). Table 3 shows the statistical analysis results for the spatial distribution of "good quality" pixels. In the comparative experiment, in addition to analyzing the "good quality" pixels in the observation data, the correlation between the altitude factor and the number of "good quality" pixels in the data was also analyzed and studied.

Analyzing the experimental results in Table 2, the LST data reconstructed by using the LPRN method show similar characteristics to the data reconstructed by the traditional method. Compared with high-altitude areas (>1500 m), low-altitude areas (such as flat bottoms and valleys) are more likely to be obscured by coverings, resulting in fewer "good quality" pixels that can be collected. It is reflected in Table 1 that the proportion of "good quality" pixels has dropped significantly. At the same time, compared with traditional reconstruction methods, in the LST data reconstructed by LPRN (loss function using NCC), the proportion of "good quality" pixels will be further increased, and the maximum increase can reach 7%.

Based on the overall analysis of the experimental results, after removing the interference of altitude factors, the percentage of "good quality" pixels that can be collected throughout the year is between 37% and 48%. Compared with the data reconstructed by the traditional method, the data reconstructed by the LPRN method can increase the proportion of "good quality" pixels by 3%–7%.

Table 3: The spatial distribution of "good quality" pixels in the daily MODIS LST obtained after dividing by altitude (each value represents the percentage of "good quality" pixels; the data reference comes from [35]).

Year	0–499 m	LPRN	500–1499 m	LPRN	>1500 m	LPRN
2000	30.0	35.3	36.0	40.2	36.1	43.2
2001	32.4	36.5	40.3	46.5	40.5	46.1
2002	27.0	30.9	33.6	40.3	35.9	39.3
2003	35.5	38.2	46.1	49.9	47.6	53.8
2004	31.5	36.7	39.2	46.2	41.6	46.9
2005	35.1	40.0	44.7	50.1	45.1	50.3
2006	34.7	39.2	43.8	44.5	46.1	51.0
2007	38.2	45.1	46.6	53.2	46.5	52.6
2008	28.3	33.1	33.4	40.1	34.0	46.2
Mean	32.5	37.2	40.4	45.6	41.5	47.7

In order to prove that the use of different loss functions affects the effect of data reconstruction, the experiment designed a comparative experiment with the loss function as a variable (only RMSE and NCC are listed here; MAE and MSE have poor results, so they are not listed here). Figure 5 shows the results of the proportion of "good quality" pixels in the LST data reconstructed by the model trained using RMSE and NCC as the loss function in an area with an altitude of less than 500 m.

In Figure 5, "ori" represents the proportion of "good quality" pixels in the output result using traditional reconstruction methods, "RMSE" represents the proportion of "good quality" pixels in the output result of using RMSE as the loss function of the LPRN model, and "NCC" represents the proportion of "good quality" pixels in the output result of using NCC as the loss function of the LPRN model. In the area below the altitude of 500 m, the proportion of "good quality" pixels in the output result of the model using the loss function NCC can be increased by about 4.7% on average. The reconstruction effect of the model based on the loss function RMSE is improved to a certain extent compared with the traditional method, but it is lower than the LST data reconstruction effect of the model based on the loss function NCC.

Figure 6 shows the results of the proportion of "good quality" pixels in the LST data reconstructed by the model trained using RMSE and NCC as the loss function in an area with an altitude ranging from 500 m to 1500 m. The proportion of "good quality" pixels in the output result of the model using the loss function NCC can be increased by about 5.2%.

Figure 7 shows the results of the proportion of "good quality" pixels in the LST data reconstructed by the model trained using RMSE and NCC as the loss function in an area with an altitude above 500 m to 1500 m. The proportion of "good quality" pixels in the output result of the model using the loss function NCC can be increased by about 6.2%.

The experimental results proved that the model effect trained by the deep learning network has great correlation with the choice of loss function. In different application scenarios, the use of different loss functions will play different roles. In the LST data reconstruction experiment, we compared the use of RMSE as the loss function with the use of NCC as the loss function, indicating that in the application scenario of LST data reconstruction, using the loss function NCC will significantly increase the proportion of "good quality" pixels, which proves that NCC is more suitable for this field.

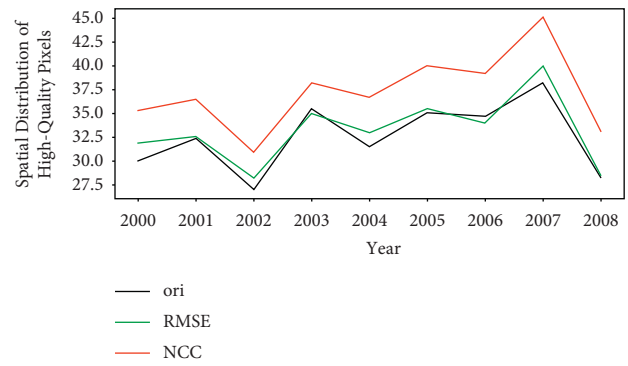


Figure 5: Proportion of "good quality" pixels at altitudes <500 m (RMSE and NCC).

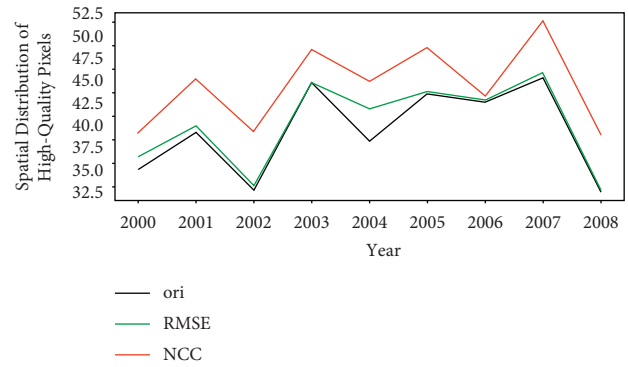


Figure 6: Proportion of "good quality" pixels at altitudes 500 m–1500 m (RMSE and NCC).

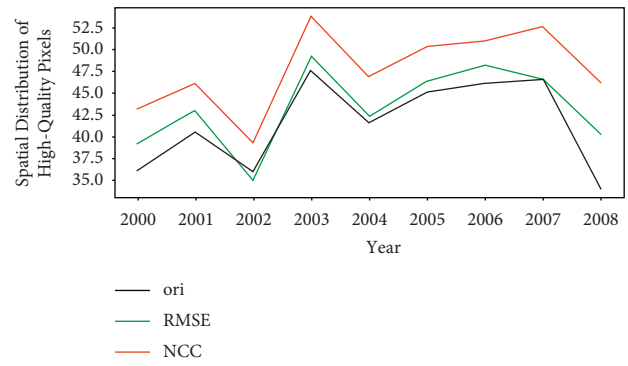


Figure 7: Proportion of "good quality" pixels at altitudes >1500 m (RMSE and NCC).

4. Conclusions

High-resolution, high-precision satellite remote sensing data have tremendous value for many fields including the Earth sciences. However, the satellite remote sensing data collected

are often contaminated, particularly by cloud cover. LST data processed using existing methods lose significant information due to occlusions, necessitating one or more passes through reconstruction methods to restore the missing information.

In this article, we use the histogram equalization and inverse distance weighting method to preprocess the LST raw data, remove the outliers in the data and construct the standard dataset usable by the deep learning network, and finally use the LPRN framework to reconstruct LST data. In the experiment to verify the accuracy of the reconstructed data, the experiment proved that the LPRN method can reduce the error between the training data and the label data. Therefore, in future extended experiments, the reconstruction effect of the LPRN method can be further enhanced by using higher precision or higher precision label data. Compared with traditional reconstruction methods, in the LST data reconstructed by LPRN, the proportion of "good quality" pixels will be further increased, and the maximum increase can reach 7%. Compared with traditional reconstruction methods, in the LST data reconstructed by LPRN, the proportion of "good quality" pixels will be further increased, and the maximum increase can reach 7%. In addition, the experimental results show that compared with high-altitude areas (>1500 m), low-altitude areas (such as flat bottoms and valleys) are more likely to be obscured by coverings, resulting in fewer "good quality" pixels that can be collected, and the proportion of "good quality" pixels will drop significantly. Finally, the experiment also found that choosing a suitable loss function has a significant effect on the experimental results. The experimental results potentially showed that increasing amounts of training data will change the experimental results. This is related to the inherent nature of the deep learning method. Therefore, it can be expected that when the amount of data reaches a certain level in the future, deep learning methods can explore more hidden information in the data.

Existing deep learning framework research has shown that training data with time continuity contain potential time and space attributes, with the ability to predict the future to some degree based on them. In the field of Earth sciences, explorations of these phenomena are rare. In the future, we plan to incorporate more diversified relevant data into the model, such as wind speed, sunshine duration, and light intensity, by integrating multiple data fusion methods into the data reconstruction methods. In this way, we hope to reduce the effects of data information loss caused by external interference and to maximize the value of data.

Data Availability

Publicly available datasets were analyzed in this study. These data can be found in the following website: https://modis.gsfc.nasa.gov/data/dataprod/mod11.php.

Disclosure

The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the National Key Research and Development Program of China (grant nos. 2020YFA0607900 and 2018YFC1507005), the National Natural Science Foundation of China (grant no. 42075137), and the Sichuan Science and Technology Program (grant no. 2020YFG0189).

References

[1] L. Chen, W. T. Huang, J. Wu et al., "Ore-forming temperature variation and prospecting target of the Wuxu ore field in Guangxi," *Geochimica*, vol. 44, no. 6, pp. 546–555, 2015.

[2] J. A. Voogt and T. R. Oke, "Thermal remote sensing of urban climates," *Remote Sensing of Environment*, vol. 86, no. 3, pp. 370–384, 2003.

[3] Z. L. Li, S. B. Duan, B. H. Tang et al., "Review of methods for land surface temperature derived from thermal infrared remotely sensed data," *Journal of remote sensing*, vol. 20, no. 5, pp. 899–920, 2016.

[4] X. Guo and G. Cheng, "Advances IN the application OF remote sensing to evapotranspiration research," *Advances in Earth Science*, vol. 19, no. 1, pp. 107–114, 2004.

[5] L. H. Zhang, Y. N. Chen, R. F. Zhao, and W. H. Li, "Impact OF temperature and soil water content ON soil res-piration IN temperate deserts, China," *Chinese Journal of Plant Ecology*, vol. 33, no. 5, p. 936, 2009.

[6] A. J. Arnfield, "Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island," *International Journal of Climatology*, vol. 23, no. 1, pp. 1–26, 2003.

[7] J. D. Kalma, T. R. McVicar, and M. F. McCabe, "Estimating land surface evaporation: a review of methods using remotely sensed surface temperature data," *Surveys in Geophysics*, vol. 29, no. 4, pp. 421–469, 2008.

[8] F. N. Kogan, "Operational space technology for global vegetation assessment," *Bulletin of the American Meteorological Society*, vol. 82, no. 9, pp. 1949–1964, 2001.

[9] Y. Jiang, D. Yu, M. Zhao, H. Bai, C. Wang, and L. He, "Analysis of semi-supervised text clustering algorithm on marine data," *Computers, Materials & Continua*, vol. 64, no. 1, pp. 207–216, 2020.

[10] E. Valor and V. Caselles, "Mapping land surface emissivity from NDVI: application to European, African, and South American areas," *Remote Sensing of Environment*, vol. 57, no. 3, pp. 167–184, 1996.

[11] L. Zhang, L. Bai, X. Zhang, Y. Zhang, F. Sun, and C. Chen, "Comparative variance and multiple imputation used for missing values in land price DataSet," *Computers, Materials & Continua*, vol. 61, no. 3, pp. 1175–1187, 2019.

[12] Z.-L. Li, B.-H. Tang, H. Wu et al., "Satellite-derived land surface temperature: current status and perspectives," *Remote Sensing of Environment*, vol. 131, pp. 14–37, 2013.

[13] Z. Wan, Y. Zhang, Q. Zhang, and Z.-L. Li, "Quality assessment and validation of the MODIS global land surface temperature," *International Journal of Remote Sensing*, vol. 25, no. 1, pp. 261–274, 2004.

[14] S.-B. Duan and Z.-L. Li, "Spatial downscaling of MODIS land surface temperatures using geographically weighted regression: case study in northern China," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6458–6469, 2016.

[15] H. Shen, X. Li, Q. Cheng et al., "Missing information reconstruction of remote sensing data: a technical review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 61–85, 2015.

[16] S. Xu, J. Cheng, and Q. Zhang, "Reconstructing all-weather land surface temperature using the bayesian maximum entropy method over the Tibetan plateau and Heihe river basin," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3307–3316, 2019.

[17] X. Zhang, J. Zhou, F.-M. Göttsche, W. Zhan, S. Liu, and R. Cao, "A method based on temporal component decomposition for estimating 1-km all-weather land surface temperature by merging satellite thermal infrared and passive microwave observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4670–4691, 2019.

[18] J. P. A. Martins, I. F. Trigo, N. Ghilain et al., "An all-weather land surface temperature product based on MSG/SEVIRI observations," *Remote Sensing*, vol. 11, no. 24, p. 3044, 2019.

[19] M. K.-P. Ng, Q. Yuan, L. Yan, and J. Sun, "An adaptive weighted tensor completion method for the recovery of remote sensing images with missing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3367–3381, 2017.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[21] B. Zhan, D. Li, X. Wu, J. Zhou, and Y. Wang, "Multi-modal mri image synthesis via gan with multi-scale gate mergence," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[22] Y. Wang, L. Zhou, B. Yu et al., "3D auto-context-based locality adaptive multi-modality GANs for PET synthesis," *IEEE Transactions on Medical Imaging*, vol. 38, no. 6, pp. 1328–1339, 2018.

[23] Y. LeCun, B. Boser, J. Denker et al., "Handwritten digit recognition with a back-propagation network," *Advances in Neural Information Processing Systems*, vol. 2, 1989.

[24] M. Chen, X. Luo, Y. Zhu, Y. Li, W. Zhao, and J. Wu, "An apriori-based learning scheme towards intelligent mining of association rules for geological big data," *Intelligent Automation and Soft Computing*, vol. 26, no. 5, pp. 973–987, 2020.

[25] W. Xie and Y. Li, "Hyperspectral imagery denoising by deep learning with trainable nonlinearity function," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 1963–1967, 2017.

[26] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1795–1799, 2017.

[27] M. Shah, R. U. Qureshi, N. G. Khan, M. Ehsan, and J. Yan, "Artificial Neural Network based thermal anomalies associated with earthquakes in Pakistan from MODIS LST," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 215, Article ID 105568, 2021.

[28] Z. Wan, *MODIS Land Surface Temperature Products Users' Guide*, Institute for Computational Earth System Science, University of California, Santa Barbara, CA, USA, 2006.

[29] Z. Wan, *Collection-5 MODIS Land Surface Temperature Products Users' Guide*, ICESS, University of California, Santa Barbara, CA, USA, 2007.

[30] P. Rakwatin, W. Takeuchi, and Y. Yasuoka, "Restoration of Aqua MODIS band 6 using histogram matching and local least squares fitting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 2, pp. 613–627, 2008.

[31] B. Chen, B. Huang, L. Chen, and B. Xu, "Spatially and temporally weighted regression: a novel method to produce continuous cloud-free Landsat imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 1, pp. 27–37, 2016.

[32] P. Scaramuzza and J. Barsi, "Landsat 7 scan line corrector-off gap-filled product development," *Proceeding of Pecora*, vol. 16, pp. 23–27, 2005.

[33] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.

[34] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, pp. 234–241, Springer, Berlin, Germany, 2015.

[35] M. Neteler, "Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data," *Remote Sensing*, vol. 2, no. 1, pp. 333–351, 2017.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 8494410, 10 pages
https://doi.org/10.1155/2021/8494410

Research Article

An Empirical Study on Virtual English Teaching System Based on the Microservice Architecture with Wireless Internet Sensor Network

Gaimin Jin,[1] **Lifang He**,[1] **and Sang-Bing Tsai**[2]

[1]*Shi Jia Zhuang University of Applied Technology, Hebei 050081, China*
[2]*Regional Green Economy Development Research Center, School of Business, Wuyi University, Nanping, China*

Correspondence should be addressed to Gaimin Jin; schoen_jin@126.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 14 July 2021; Revised 2 August 2021; Accepted 25 August 2021; Published 2 September 2021

Academic Editor: Xianyong Li

We design in this paper a virtual English teaching system based on a wireless sensor network using the features of microservice architecture such as low coupling, lightweight, and high autonomy. This paper explains the characteristics of microservice architecture, uses the wireless sensor network as a carrier, uses the technology provided by Spring Cloud ecology to practice microservices, and makes a specific design and implementation of microservice architecture. The system allows teachers and students to upload and download English course resources, adopts the knowledge map to classify and manage course resources, realizes the diversified search and sorting function of course resources, greatly facilitates users to query and manage course resources, and improves the utilization rate. The internal functional modules of the platform are reasonably divided with the microservice design principle and are developed and deployed independently. For the security of the system, the authentication mechanism of the device and the authentication and authentication strategy of the application interface are designed, respectively. Finally, functional and nonfunctional tests were conducted on the system through the built test environment. The test results show that the virtual English teaching system designed and implemented in this paper has good feasibility and scalability. Through the research of this paper, it brings great help to the wireless sensor network virtual English teaching system and also improves the learning efficiency of students.

1. Introduction

With the development of Internet technology, the theory of microservices is becoming more and more mature, and the related technologies around microservice architecture are gradually improved. The development cost of microservices is greatly reduced, and many traditional monolithic application architectures have started to transform to microservice architecture [1]. The reason for this is that the increasing number of users and types of access has increased the difficulty and cost of system development, and further development under the previous architecture is much more difficult than splitting into the microservice architecture. Most of the systems have been out of this status quo, the old system cannot adapt to the new technology, and the development is also slow [2, 3]. Some others still use the overall architecture, but there is a constant demand for increased functionality, there are also problems brought about by the increase in users, the system begins to become chaotic, the subsequent development of the software is slow, the system is full of loopholes, various problems are constantly exposed, and the accumulation is difficult to return. Traditional English teaching faces infinite impacts and challenges while welcoming the new environment and opportunities of the computer network environment: how to guide the efficient integration of the two resources, that is, to give full play to the advantages of the information age. The computer network adapts to the needs of modern education development and truly integrates into the curriculum as an organic part. It has become an important issue to promote the transformation of English teaching from traditional to modern and also constitutes the focus of the current research on the reform of virtual English teaching.

The wireless sensor network virtual English teaching system has largely improved the teaching mode and enriched the teaching methods. Wireless sensor network virtual English teaching system makes full use of the characteristics of network resource sharing, solving the impact of time and space on traditional experiments, while greatly reducing teaching costs and improving teaching efficiency. Particularly for some more dangerous experiments, the role of wireless sensor network virtual English teaching system is particularly prominent due to the fact that the limited energy of wireless sensor network nodes, limited computing power, the variability of the geographical environment, and the mobility of nodes resulting in limited communication between nodes and other problems make it more difficult to carry out experiments in the teaching process. To build a virtual English teaching system for wireless sensor networks to help students master the experimental content better and faster and use their precious time for critical learning content, the wireless sensor network virtual English teaching system involves several directions of computer technology, which is developed by simulation and computer network and combined with database [4]. There are abundant virtual software resources on the network to provide good conditions for the construction of the wireless sensor network virtual English teaching system. Students can run the wireless sensor network virtual English teaching system by configuring the parameters on the PC, which is good for cultivating students' learning interests. Therefore, it is of good practical significance to research and develop a wireless sensor network virtual English teaching system with good generality and good simulation performance [5, 6].

The main work of this paper designs and implements a wireless sensor network virtual English teaching system. By analyzing the current problems of backward teaching methods, insufficient vivid teaching contents, confusing teaching resource management, poor teaching quality, and weak teacher-student interaction in the process of English teaching, we design and develop a wireless sensor network virtual English teaching system that meets the actual English teaching, and the system effectively improves the teaching efficiency and teaching quality of English courses. The system effectively improves the efficiency and quality of English teaching [7]. The first section discusses the background and significance of this research paper in detail and discusses the main work and section arrangement. Section 2 introduces the related work of this research and analyzes the current status of the research. Section 3 investigates the requirements and network model of the virtual English teaching system based on microservice architecture for wireless sensor network, which lays the foundation for the system design and development, and also provides a detailed design of the virtual English teaching system for wireless sensor network. Section 4 focuses on the relevant analysis of this paper's research, and the value of this paper's research is demonstrated through experimental analysis. Section 5 concludes and summarizes the main work done in this paper and explains the future system improvement based on the current work status.

2. Related Work

With the rapid development of the Internet industry and intensified competition, how to quickly complete online product iterations and enrich product functions has become the key to increasing active users of Internet products, and a development method called microservices has been favored by major Internet companies at this time and has become the standard architecture for Internet product development. Graham argues that virtual English teaching by wireless sensor network is a combination of the current learners' personalized learning needs [8]. Mo et al. proposed a wireless sensor network virtual English teaching with microservice architecture, and the system first analyzes and matches learners' personalities using Myers-Briggs Type Indicator (MBTI) [9]. Melis et al. used the Bruce learning model to provide matching learning contents for different learners, and the teaching model is experimentally verified to improve students' learning efficiency. Management and service are the main problems that solve future classroom teaching [7]. Winkowska et al. proposed a virtual English teaching system combining an online teaching management system and teaching resource platform to optimize the existing teaching methods and improve the efficiency of school teaching resource management [10].

Since microservice architecture is broadly defined as the use of a set of small services to design service-oriented software, there is a certain application complexity between independently deployed computing units in the microservice architecture, and such complexity can lead to security vulnerabilities [11]. Liu et al. proposed the microservices-based data service framework, which overcomes the problems of large volume and complex protocols of traditional data service platforms and at the same time greatly improves the maintainability and scalability of software to achieve the purpose of decoupling and decentralization [12]. Chen et al. proposed a data-driven learning concept based on microservice architecture learning and the English learning approach and related teaching models driven by this concept can be seen as development and extension of the previous resource-based learning models [13]. By studying the interactive teaching mode of English listening and speaking classes, Bao et al. argued for the positive role of multimedia teaching in improving college students' English listening and speaking skills and applied information technology to reading teaching to explore a new model for transforming English reading teaching in colleges and universities [14].

Through combining and analyzing the research of virtual English teaching mode in the computer network environment at home and abroad and the wireless sensor network of microservice architecture, we can find that the research where researchers put more energy mainly focuses on the theory of virtual English teaching in the computer network environment. In terms of exploration, optimization of teaching modes for different course types, and exploration of teaching strategies, there is a very little research on system performance. There have been a lot of excellent research results, but there are still some problems under the

prosperous scene, such as individual studies only rigidly apply the theories of English teaching in the traditional environment to study the teaching practice in the new environment [15]. The author believes that these problems should be avoided in future research; otherwise, the reliability and validity of the new round of research will be seriously affected [16]. Firstly, according to the problems of confusing English teaching resource management and complicated types of teaching courses, a comprehensive wireless sensor network virtual English teaching system is designed and implemented so that teachers and students can manage English course resources through this module, such as uploading, deleting, editing, and other operations, and students can view and download English course resources through this module and other functions, and secondly, the module also classifies and manages the course resources in the way of knowledge map, realizes the diversified search and sorting function of course resources, greatly facilitates users to query and manage the course resources, and improves the efficiency of English course resource management. The wireless sensor network virtual English teaching system based on microservice architecture facilitates students to record their English learning and enrich the English teaching course content to improve the practicality of the system.

3. Research on Virtual English Teaching System Based on Microservice Architecture for Wireless Sensor Network

3.1. Research on the Needs of Virtual English Teaching. Virtual English is a framework and procedure constructed under the guidance of a specific teaching concept that can reflect the law of teaching development and has constructive characteristics such as clear purpose, stable structure, and strong practicality. The decolonization of virtual English is the healthy development goal of modern virtual English, which is gradually explored and formed in the teaching practice of ecological classrooms, and its construction can in turn promote the ecological development of classroom teaching. The ecological model is guided by the theory of educational ecology and the concept of "student-centered" education, designs various forms of teaching activities such as multidimensional interaction according to students' ability levels and knowledge needs, and promotes students' natural, harmonious, and free growth and development under good artificial conditions by balancing the ecological position of several teaching components [17]. This virtual English application in the classroom is a very important tool for teaching and learning. The application of virtual English in classroom teaching is not only to realize the reasonable arrangement of a single virtual English but also to study the complementary adaptation and coexistence between multiple virtual English, to select the suitable virtual English for different student groups, different teaching contents, and different teaching environments to realize the complementary advantages of teaching resources.

To make the system have better module reusability and technical advancement, the microservice architecture-based wireless sensor network virtual English teaching system mainly adopts the microservice framework for layered design, dividing each functional module into three layers of business logic layer, model layer, and view layer for design, and each layer is responsible for different system functions. The Spring framework is mainly responsible for the design of the business logic layer, including the system login and registration business logic, English course resources sorting/querying/searching logic, English course test logic, English test preparation logic, and authentication logic [18, 19]. These logic module interfaces can provide a unified standardized logic function for the view and model layers; the Hibernate model layer is mainly responsible for system background data management, such as data addition, deletion, data table creation, deletion, and database backup. The background data of the hybrid English teaching management system mainly include basic student information, English course resources, English test resources, and English teaching resources and provide data services for the upper layer through a unified data interface. The view layer in the Struts framework is mainly responsible for system interface customization and display, such as the main interface of the system, user personal center interface, English course resource management interface, English course test interface, English teaching interface, and other modules. The architecture of each layer of the virtual English teaching system based on the microservice framework is shown in Figure 1.

The system performance requirement analysis mainly analyzes the system hardware and software environment, system quality, system external interfaces, and so on. Specifically, the system performance requirement analysis mainly includes system correctness, stability, security, portability, system response speed, server response speed, memory usage, and CPU utilization rate. The microservice architecture-based wireless sensor network virtual English teaching system is based on microservice architecture, using microservice architecture to implement each module of the system. Through the Internet network using multimedia technology for online English teaching, the system requires a good performance index. In this paper, we mainly analyze the system reliability, hardware and software environment, and performance indicators. Among them, the system reliability requirements are analyzed as shown in Table 1.

Feasibility analysis mainly refers to the feasibility analysis from system development technology, system resources required, and system market prospect before the system is designed and developed to ensure that the system can be carried out smoothly in the design and development process

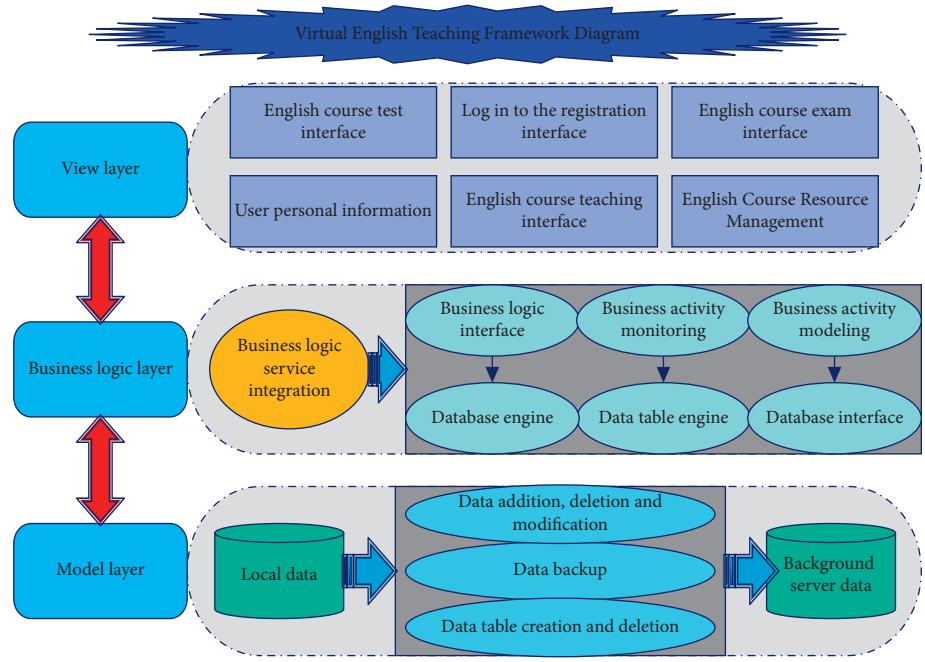


Figure 1: Virtual English teaching system architecture diagram.

Table 1: System reliability requirements.

Reliability requirements	Reliability detailed requirements	Reliability index
Correctness	The cumulative error rate of each functional module of the system	Less than 0.5%
Stability	The cumulative error rate of continuous system operation	Less than 0.5%
Restorative	The system shuts down and restarts abnormally, and the function can operate normally	Higher than 99.5%
Safety	System data security, zero loss	Higher than 99.9%

[19]. The feasibility analysis of the wireless sensor network virtual English teaching system based on microservice architecture is mainly from three aspects: technical feasibility, resource feasibility, and economic feasibility.

3.2. Microservice Architecture Wireless Sensor Network Model Construction. In wireless sensor networks, the irregularity of network environment and data collection requires the controller to have strong adaptive capability [20]. In this paper, we propose to combine fuzzy control with traditional PID controller and introduce it into wireless sensor network to realize its parameter self-tuning, calculate the instantaneous queue length and real-time packet loss rate of message queue, and realize adaptive congestion control.

To ensure that the instantaneous queue length of the message queue within the node has absolute safety and reliability, the setting of the message discard probability will be chosen to find the maximum value of the message discard probability. Using $f(x)$ as the objective function of the cuckoo search, the maximum value of the message discard probability $f(x)$ is calculated by continuous update iterations using the jump path and step size of the $g(m, n)$ flight update search when the PID parameters change. The path update expression of cuckoo search is shown in equation (1), where $f(x, n)$ denotes the position of the message discard probability at this k-th update at this n-th iteration, α denotes the step size factor, which obeys normal distribution, and $g(m, n)$ denotes Levy flight.

$$f(x,n) = \{f(x,n-1) + \beta\} \forall g(m,n). \tag{1}$$

A random wander is performed by Levy flight to obtain the location of the next updated nest, and the expression is as in the following equation:

$$g(m,n) = \frac{\mu\xi(\mu)\cos(\mu\tau/2)}{\tau} * \frac{1}{\sum_{i=1}^{n} \mu^{m-1}}, \quad m \subseteq [1, +\infty]. \tag{2}$$

The parameters such as initial population size and a maximum number of iterations M are initialized. In the simulated scenario, 200 random network nodes are set, so the initial population size $N = 200$ and the maximum number of iterations $M = 200$ are set, and the probability of discovery is as in the following equation:

$$f(n) = \frac{\sum_{i=1}^{n} f(i,t)}{t} * 100\%. \tag{3}$$

The objective function is set to the message discard probability $f(x)$, at which point, the expression for the message discard probability within a node is updated to equation (4). In this algorithm, the maximum number of

iterations computed is H, where there are no nested loops, so the computational complexity spent in each iteration to compute the optimal solution does not exceed $O(H)$. The maximum population size is G. Therefore, the space complexity of the cuckoo search algorithm is $O(G)$.

$$f(x) = H(f(0)) + \exists H(f(1)) + \frac{1}{\sum_i^n (G(i) - \exists G(i-1))}. \tag{4}$$

The integrated model of the virtual English teaching system reflects a complete and sound pattern in the system view. Not only are the four step-by-step design elements of teaching objectives, learning starting points, process methods, and outcome evaluation interconnected to ensure the functioning of the whole teaching system, but also each of the four elements has a unique structure within itself to achieve specific teaching functions. The wireless sensor network model establishes a set of vocabulary acquisition level objectives reflecting the principles of English subject pedagogy and a corresponding evaluation system combining subjective and objective criteria. The two parties are responsible for different types of teaching tasks in different contexts, and finally, the theoretical requirements of individualized teaching are fully implemented through a synchronized and integrated teaching process and teaching method strategy.

Given the search direction (x, t), the step size can be calculated by being. In the backtracking line search, the step size is taken as $s = f(x, \beta)$, and the update rule for t is shown in equation (5), which uses the step size as a rough measure of the proximity to the central path. Generally, $\beta = 2$ and $s = 0.05$.

$$H(t) = \max\left\{\max\{\beta \min(3m/\tau, t), t\}, \sum_{i=1}^{n} f(i) * \min(s_i, s_{i-1})\right\}. \tag{5}$$

In local routing-based topology, the multihop mode is used to save energy and achieve efficient routing. Sensor nodes collect information from the surrounding environment to form a data-centric forwarding route. Building a cluster can also be divided into four phases, including the election phase of the cluster head node, the broadcast phase of the cluster head node, the establishment phase of the cluster, and the generation phase of the scheduling mechanism within the cluster. The cluster scheduling mechanism means that the cluster head assigns time slots to the cluster members that allow them to send data using the Time Division Multiple Access (TDMA) method. Here, the cluster head can choose to consolidate the intracluster messages before forwarding them out. This is followed by the data stabilization phase, where messages are sent according to the currently established clusters. After some time, the stabilization phase ends and a new cluster establishment cycle is entered.

$$G(t) = \frac{f(t)}{1 - f(t) * |\tau| \sin(1/f(t))|}. \tag{6}$$

3.3. Virtual English Teaching System Design. To develop a system using microservice architecture, it is necessary to establish the infrastructure, and at the same time, it is necessary to be able to divide the services according to the requirements, so that each service module can be developed and run independently [21–23]. Therefore, this virtual lab platform uses a combination of microservices and layered architecture. With the API gateway as the boundary, the front end and back end of the system are separated. RESTful requests are used to interact the web pages with the back-end services. To simplify the call logic, the back-end service provides an API gateway as the call interface for the web front end to access the back end. As an important part of the microservices architecture, the API gateway aggregates the invocation logic of multiple microservices, reducing the number of client requests and optimizing the client experience. In the microservice architecture, each different back-end service may have a user-maintained business, so there will be a lot of redundant login verification and signature verification in microservices. We can separate these functions in microservices into a separate OAuth2 authentication authorization server and then call the OAuth2 server through the API gateway to filter user requests and ensure the security of the service is guaranteed (see Figure 2).

The wireless sensor network virtual English teaching system based on microservice architecture mainly mixes a variety of multimedia methods, such as pictures, text, audio, video, and other contents, using the advantages of network teaching. School students can learn English courses online and provide comprehensive guidance to students on English words, English grammar, English speaking, reading comprehension, English composition, English translation, and other contents. The system also provides English course mock exams and English test questions to allow students to customize their learning plans according to their conditions and complete English learning tasks in a targeted manner, providing students with learning efficiency. The main functional modules of the wireless sensor network virtual English teaching system based on microservice architecture are shown in Figure 2.

In the system, the logical structure of the database is generally independent of the specific functions and is generally more abstract. The design of the system database logical structure can generally adopt three different design methods: the top-down design method firstly designs the overall data structure and then gradually enriches and refines the data structure from top to bottom; the bottom-up design method firstly designs the specific data structure and then gradually merges and unifies the data types from bottom to top; the expansion-by-expansion design method firstly designs the core The bottom-up design approach

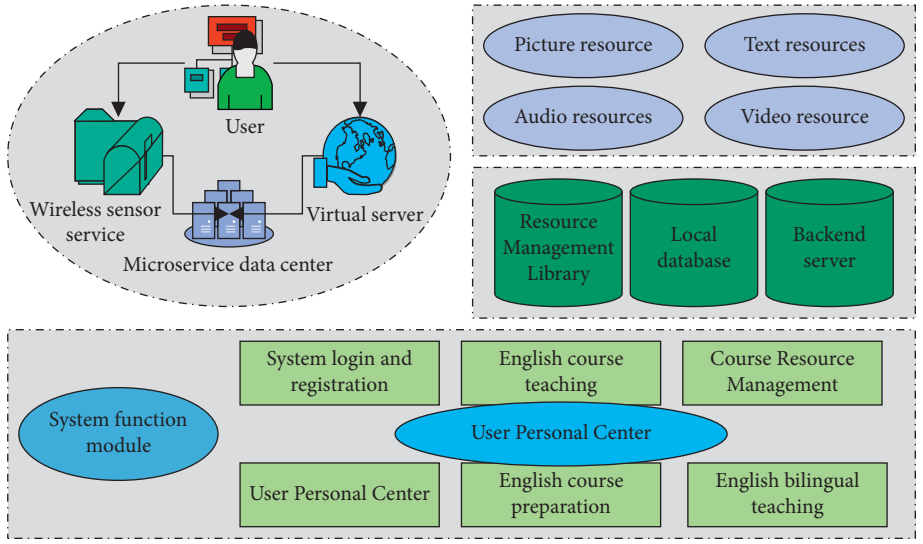


Figure 2: Main functional modules of virtual English teaching system.

starts with the design of the core data structure and continues to expand outward until the design of all data structures is completed. According to the actual data management requirements and data structure characteristics of the English teaching interactive management system designed and developed in this project, a combination of top-down and bottom-up database logical structure design methods was adopted in the design process.

4. Analysis of Results

4.1. Network Model Analysis. This section compares the average throughput of three congestion control algorithms when the number of nodes increases. Throughput is a direct indicator of network performance. By increasing the number of nodes, the amount of data passed by messages is indirectly increased to observe the variation of throughput in more complex scenarios. Throughput is limited by the computing power of the hardware device itself, so the magnitude of its magnitude does not vary too significantly. The throughput comparison curves of the three are shown in Figure 3. As can be seen from Figure 3, when the number of nodes increases, the number of messages for data transmission will increase, the time for data transmission between nodes will shorten, and the throughput will increase accordingly. The throughputs of all three congestion control algorithms show different degrees of increase when the number of nodes increases. The CFPID algorithm has the highest throughput when the number of nodes is small. As the number of nodes grows, the growth in throughput of IBLUE and PID is gradually surpassed by CFPID, reflecting the superiority of CFPID in terms of later stability and adaptive adjustment in complex situations (see Figure 3).

We compare the real-time packet loss rate of nodes under the regulation of three congestion control algorithms. When the distribution density of nodes rises, the overall total amount of data in the network increases and the number of

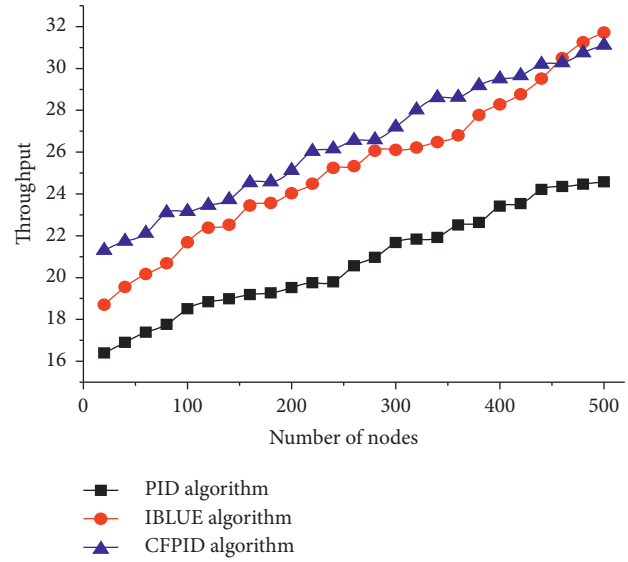


Figure 3: Throughput variation comparison curve.

packets to be transmitted in the message queue within a node will increase, resulting in an elongated transient queue length. When the queue length exceeds the desired value, it will intensify the network congestion, then the packet drop operation is required to control the queue length by dropping a certain number of packets through the message drop probability, and the packet drop rate will change with the change of the instantaneous queue length to control the queue length around the desired value. But the packet loss rate keeps changing, which will make the queue length of the message queue keep changing, which leads to the instability of the network, and the fluctuation of the network will also indirectly affect the overall network working status. Therefore, the convergence speed and stability of the instantaneous packet loss rate will determine the performance of the network. In the experiments of this section, the initial number of nodes is set to 100, and the variation curve of the

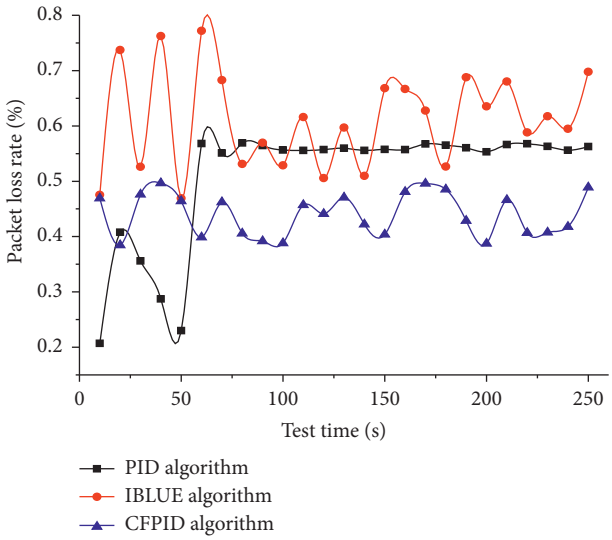


Figure 4: Real-time packet loss rate variation curve.

real-time packet loss rate of network nodes is shown in Figure 4.

As can be seen in Figure 4, when the distribution of network nodes is denser, the elevated data volume and the shortened transmission time lead to a growing message queue of nodes. When the message queue is stretched significantly, the CFPID congestion control algorithm can detect the danger more accurately and quickly, initially use the maximum packet loss rate to limit the further growth of the queue length, and pull the instantaneous queue length with the trend of congestion back to the controllable range quickly to suppress the aggravation of congestion. After a small amount of oscillation, the equilibrium point will be found as quickly as possible and stabilized in a region with less variation, and then only small adjustments are needed to keep the instantaneous queue length in a relatively stable and healthy state (see Figure 4).

4.2. System Performance Analysis. Figure 5 shows the message delivery rate values obtained for LBR1, LBR2, AODV, and RASeR. lBR1 (Figure 5(a)) shows the best performance, followed by LBR2, which is negatively affected by the high migration rate. aODV (Figure 5(c)) presents a low delivery rate in general and is poorly scalable. On the other hand, RASeR provides a decent messaging rate of up to 32 nodes, obtaining nearly 60%. In this protocol, we observe problems related to messaging duplication, leading to network congestion. Intuitively, the global TDMA imposes an upper limit on the number of messages sent per cycle. Therefore, if the number of received messages is continuously higher than this upper limit, it will eventually lead to a buffer overflow. This situation can be represented by the total number of messages per protocol used to route multihop traffic to the receiver, a parameter that has a significant impact on energy consumption. Although LBR1, LBR2, and RASeR share routing criteria based on hop count, RASeR is very inefficient in terms of the number of messages it generates (see Figure 5).

To ensure that the system can respond to users' requests on time when the system is put into use with concurrent access by multiple users, the Load Runner test tool was used to simulate concurrent access to the system by a large number of users and to test the response time of the system. According to the actual application of the system, concurrent access of 300 and 500 people was simulated by Load Runner, and the performance of the system was verified to meet the expected design requirements through testing. The system performance test shows that when 300 and 500 people access the system at the same time, the access time of the system can be guaranteed within 5 seconds, which fully meets the expected design requirements of the system, indicating that the system can handle concurrent access by multiple users and meet the expected design goals. The specific results of the system performance test, such as stress test and system response time, are shown in Figure 6.

There are still problems in the design and implementation of the system, the system uses the microservice architecture for English system implementation, the system operation efficiency is not high enough, and the level of intelligence still needs to be improved, so for today's intelligent Internet era, intelligent Internet has become the mainstream of the development of today's era. It is necessary to develop an English teaching management system based on artificial intelligence and machine learning, to better meet the needs of today's era. Secondly, the English resources of the system are not rich enough, and the English courseware materials are not ideal enough to cover all the English teaching tutorials, so it is necessary to further improve the English resources at a later stage to improve the practicality of the English teaching management system. It does not have grouping, group member identification, and group voice function, which cannot fully meet the teaching needs of teachers and students in the process of teaching English courses.

4.3. System Application Analysis. The purpose of this construction method is to enhance the observability and comparability of changes in vocabulary acquisition levels by maintaining the equivalence between the two tests; the test itself remains unchanged in other aspects; that is, the type of questions is in the form of English-Chinese translation, the number of items is 82, and the score is 100. The full score was 100, and the test duration was 30 minutes. The posttest data were statistically compiled and showed that the average score of the experimental group was 57.9 and the average score of the control group was 54.3. To further evaluate the actual effectiveness of the experimental and control groups on their corresponding teaching treatments, the study selected the paired-sample t-test as the inferential statistical method and used SPSS20 as the statistical test tool to first match the pre- and posttest scores of individual subjects from the experimental and control groups and then perform a hypothesis test on the overall data. The results of the

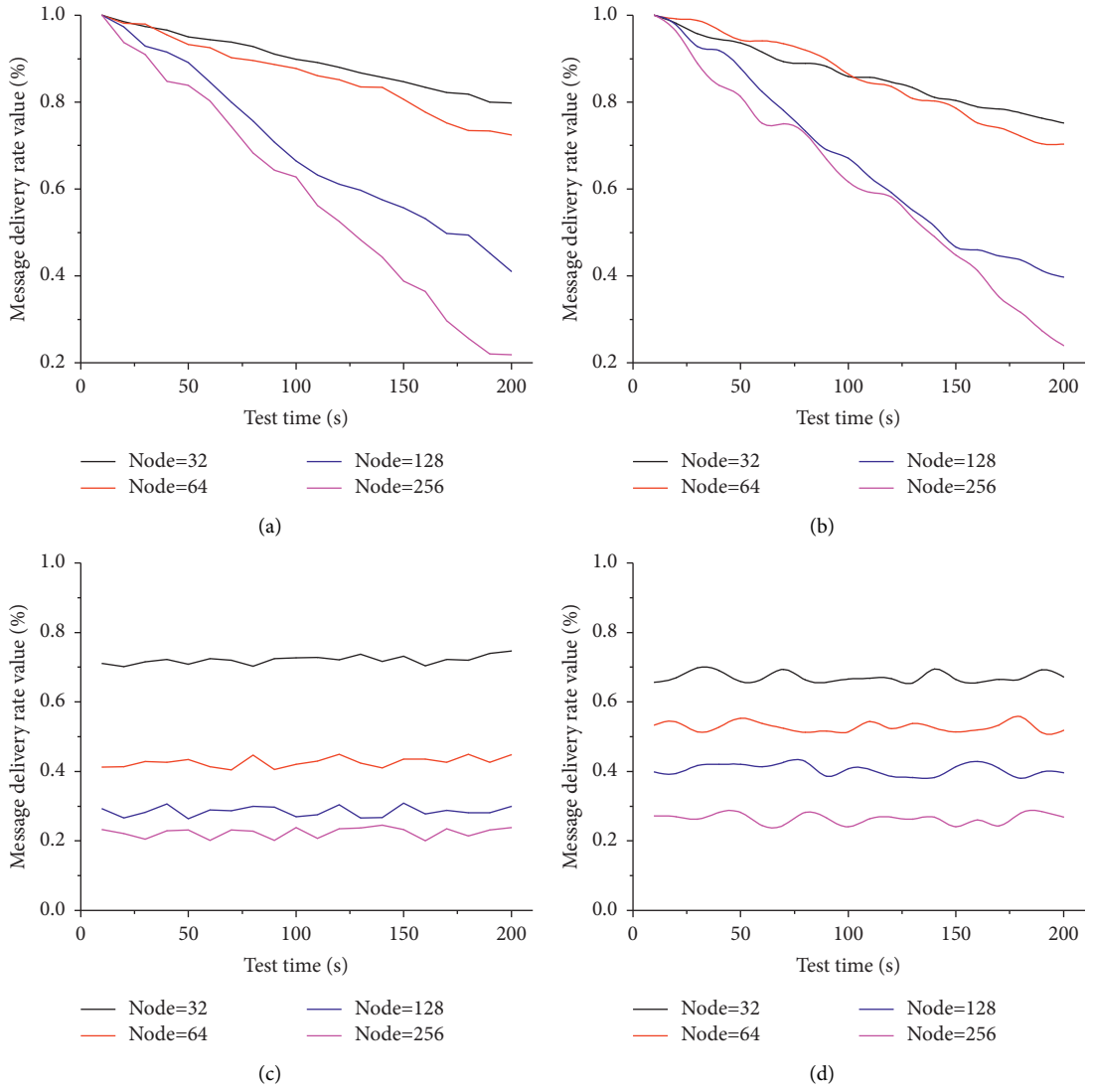


Figure 5: Number rate values for messaging. Message delivery rate value of (a) LBR1, (b) LBR2, (c) ADOV, and (d) RASeR.

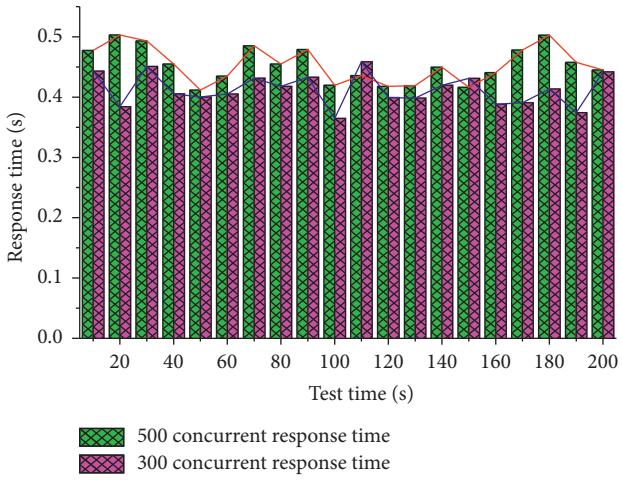


Figure 6: System processing capacity stress results.

hypothesis test on the variability of the means are shown in Figure 7.

The design goal of the virtual English teaching system is to help users teach and learn English more effectively, encapsulate the complex device access and data transmission process at the bottom of the IoT, and reduce the learning curve and usage difficulty of users. Therefore, the usability of the system is tested in three aspects, including ease of use, user experience, and development efficiency. 100 users with different age characteristics and professional knowledge backgrounds are selected for trial according to the basic operational objectives, their usage results and operational experience are scored (each index is scored out of five), the trial results are recorded and analyzed, and the statistical results are shown in Figure 8. Through the analysis of the survey results, users of different age groups and professional backgrounds can easily use the functions provided by the system, which can effectively improve the learning efficiency

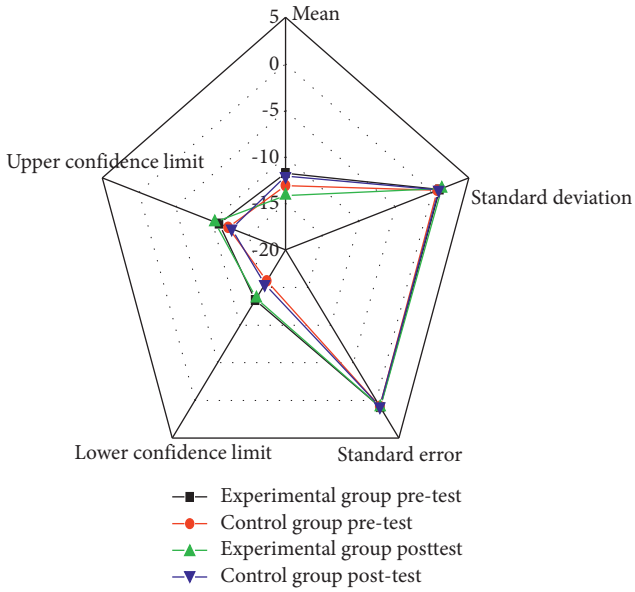


Figure 7: The *t*-test results of paired groups of experimental and control groups at their respective levels.

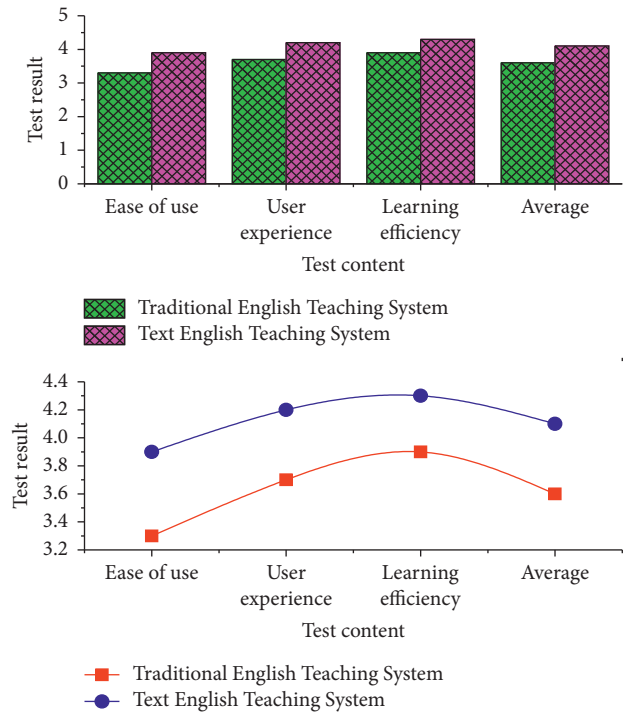


Figure 8: Availability survey results.

and reduce the learning curve of users compared with traditional teaching methods, and combined with the simple and easy-to-use functional interface, users can quickly realize the management of the device side (see Figure 8).

Through the random distribution of multiple movable network nodes, the wireless sensor network performs real-time data monitoring, adjustment, and control of the area to be measured, realizes people's data collection and control of super-large-scale scenarios, and makes a great contribution to the development of the Internet of Things. However, advanced wireless network technology has brought many value-added services and has also caused a series of data transmission reliability problems. To ensure the reliability of data transmission and avoid excessive network energy consumption and network congestion from affecting data forwarding, scientific data forwarding routing and congestion control mechanisms have become the top priority for wireless sensor networks to achieve highly reliable data transmission. Implement and test the English teaching management system, focusing on the system login and registration module, the system main interface module, the English course resource management module, the English course teaching management module, the English course testing module, the English course preparation module, the user personal center module, and so on. The realization and analysis are carried out, and the core realization codes and effect diagrams of each functional module are given. The important modules of the system are systematically tested and the test results are analyzed, which meet the needs of virtual English course teaching.

5. Conclusion

This paper initially analyzes the application prospect of the wireless sensor network, discusses the importance of this technology and the difficulty of learning this technique, introduces for this difficulty the background of the research of virtual English teaching system of the wireless sensor network, combines it with the concept of microservices, and discusses the design and implementation of a virtual English teaching system of wireless sensor network based on microservice architecture. This paper introduces the technology related to this wireless sensor network virtual English teaching system, the system requirements. In response to these requirements, the microservice architecture is used to divide it into corresponding modules and the key parts of the modules are analyzed and designed. This paper designs and implements a virtual English teaching system based on microservice architecture for wireless sensor networks, which is a more detailed design and preliminary implementation of the virtual English teaching system for wireless sensor networks, filling the gap in the research of virtual English teaching system. At the same time, the implementation of the English management system can provide quality learning support services for the majority of students, greatly stimulating students' enthusiasm for learning and improving the quality of English Teaching quality. With the arrival of the era of big data and the era of intelligence, software development has a breakthrough and development point; therefore, the future English teaching management system also needs to follow the trend of the times, using today's advanced data mining technology, machine learning knowledge, and artificial intelligence technology to do further transformation and upgrading of the system, which is one of the key concerns of the later system transformation and upgrading. Through the introduction

of new technologies, the operating speed and experience ability of the system can be better improved.

Data Availability

The data used to support the findings of this study are available within the article.

Conflicts of Interest

No conflicts of interest exist concerning this study.

References

[1] K. Siva Vignesh, "Trading platform powered by blockchain," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 11, pp. 3506–3514, 2021.

[2] A. Maatoug, G. Belalem, and S. Mahmoudi, "Fog computing framework for location-based energy management in smart buildings," *Multiagent and Grid Systems*, vol. 15, no. 1, pp. 39–56, 2019.

[3] S. S. Gill and R. Buyya, "Failure management for reliable Cloud computing: a taxonomy, model, and future directions," *Computing in Science & Engineering*, vol. 22, no. 3, pp. 52–63, 2018.

[4] Y. Simmhan, P. Ravindra, S. Chaturvedi, M. Hegde, and R. Ballamajalu, "Towards a data-driven IoT software architecture for smart city utilities," *Software: Practice and Experience*, vol. 48, no. 7, pp. 1390–1416, 2018.

[5] F. Ortin and D. O'Shea, "Towards an easily programmable IoT framework based on micro services," *JSW*, vol. 13, no. 2, pp. 90–102, 2018.

[6] M. Ling, M. J. Esfahani, H. Akbari, and A. Foroughi, "Effects of residence time and heating rate on gasification of petroleum residue," *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1837–1840, 2016.

[7] A. Melis, S. Mirri, C. Prandi, M. Prandini, P. Salomoni, and F. Callegati, "Integrating personalized and accessible itineraries in MaaS ecosystems through microservices," *Mobile Networks and Applications*, vol. 23, no. 1, pp. 167–176, 2018.

[8] H. Ma and S.-B. Tsai, "Design of research on performance of a new iridium coordination compound for the detection of Hg2+," *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, p. 1232, 2017.

[9] L. Mo, W. Sun, S. Jiang et al., "Removal of colloidal precipitation plugging with high-power ultrasound," *Ultrasonics Sonochemistry*, vol. 69, p. 105259, 2020.

[10] J. Winkowska, D. Szpilko, and S. Pejić, "Smart city concept in the light of the literature review," *Engineering Management in Production and Services*, vol. 11, no. 2, pp. 70–86, 2019.

[11] D. Bai, J. Tang, G. Lu, Z. Zhu, T. Liu, and J. Fang, "The design and application of landslide monitoring and early warning system based on microservice architecture," *Geomatics, Natural Hazards and Risk*, vol. 11, no. 1, pp. 928–948, 2020.

[12] Q. Liu, G. Lu, J. Huang, and D. Bai, "Development of tunnel intelligent monitoring and early warning system based on micro-service architecture: the case of AnPing tunnel," *Geomatics, Natural Hazards and Risk*, vol. 11, no. 1, pp. 1404–1425, 2020.

[13] G. Chen, P. Wang, B. Feng, Y. Li, and D. Liu, "The framework design of smart factory in discrete manufacturing industry based on cyber-physical system," *International Journal of Computer Integrated Manufacturing*, vol. 33, no. 1, pp. 79–101, 2020.

[14] L. Bao, C. Wu, X. Bu, N. Ren, and M. Shen, "Performance modeling and workflow scheduling of microservice-based applications in clouds," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 9, pp. 2114–2129, 2019.

[15] D. Xu and H. Ma, "Degradation of rhodamine B in water by ultrasound-assisted TiO2 photocatalysis," *Journal of Cleaner Production*, vol. 313, p. 127758, 2021.

[16] D. Gao, Y. Liu, and Z. Guo, "A study on optimization of CBM water drainage by well-test deconvolution in the early development stage," *Water*, vol. 10, no. 7, 2018.

[17] J. Xie and H. Ma, "Application of improved APO algorithm in vulnerability assessment and reconstruction of microgrid," *IOP Conference Series: Earth and Environmental Science*, vol. 108, no. 5, Article ID 052109, 2018.

[18] T. Cerny, "Aspect-oriented challenges in system integration with microservices, SOA and IoT," *Enterprise Information Systems*, vol. 13, no. 4, pp. 467–489, 2019.

[19] X. Chen, X. G. Yue, and R. Li, "Design and application of an improved genetic algorithm to a class scheduling system[J]," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 1, pp. 44–59, 2021.

[20] R. Minerva, G. M. Lee, and N. Crespi, "Digital twin in the IoT context: a survey on technical features, scenarios, and architectural models," *Proceedings of the IEEE*, vol. 108, no. 10, pp. 1785–1824, 2020.

[21] T. Grubljesic, P. S. Coelho, and J. Jaklic, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2019.

[22] L. Z. Zhang, M. Mouritsen, and J. R. Miller, "Role of perceived value in acceptance of "bring your own device" policy," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 65–82, 2019.

[23] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer gamification," *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6561730, 12 pages
https://doi.org/10.1155/2021/6561730

Research Article

Design and Implementation of Embedded Real-Time English Speech Recognition System Based on Big Data Analysis

Lifang He,[1] **Gaimin Jin**,[1] **and Sang-Bing Tsai**[2]

[1]*Shijiazhuang University of Applied Technology, Shijiazhuang, Hebei 050000, China*
[2]*Regional Green Economy Development Research Center, School of Business, WUYI University, Nanping, China*

Correspondence should be addressed to Lifang He; clare2021@126.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 14 July 2021; Revised 29 July 2021; Accepted 25 August 2021; Published 2 September 2021

Academic Editor: Xianyong Li

This article uses Field Programmable Gate Array (FPGA) as a carrier and uses IP core to form a System on Programmable Chip (SOPC) English speech recognition system. The SOPC system uses a modular hardware system design method. Except for the independent development of the hardware acceleration module and its control module, the other modules are implemented by software or IP provided by Xilinx development tools. Hardware acceleration IP adopts a top-down design method, provides parallel operation of multiple operation components, and uses pipeline technology, which speeds up data operation, so that only one operation cycle is required to obtain an operation result. In terms of recognition algorithm, a more effective training algorithm is proposed, Genetic Continuous Hidden Markov Model (GA_CHMM), which uses genetic algorithm to directly train CHMM model. It is to find the optimal model by encoding the parameter values of the CHMM and performing operations such as selection, crossover, and mutation according to the fitness function. The optimal parameter value after decoding corresponds to the CHMM model, and then the English speech recognition is performed through the CHMM algorithm. This algorithm can save a lot of training time, thereby improving the recognition rate and speed. This paper studies the optimization of embedded system software. By studying the fixed-point software algorithm and the optimization of system storage space, the real-time response speed of the system has been reduced from about 10 seconds to an average of 220 milliseconds. Through the optimization of the CHMM algorithm, the real-time performance of the system is improved again, and the average time to complete the recognition is significantly shortened. At the same time, the system can achieve a recognition rate of over 90% when the English speech vocabulary is less than 200.

1. Introduction

English speech recognition is a branch of pattern recognition, which is an interdisciplinary subject integrating microelectronics, communications, computers, automation, and acoustics [1, 2]. The most important technology of English speech recognition is the construction of speech signal processing technology and training model. The ultimate goal is to hope that humans can communicate with computers. This kind of human-computer dialogue scenes often appears in science fiction movies. In fact, this is a very complicated technology. In addition to English speech signal processing technology, how to recognize sounds and understand what is said is challenging. Due to the discontinuity of English speech, the difference in accent and pitch of each person, and the difference in speaking speed and volume, these factors will increase the difficulty of English speech recognition, so English speech recognition has always been regarded as a challenging subject. Early English speech input must separate the relationship between each word clearly and input each word separately, which is different from what we usually say. The object to be recognized is also the recognition of a specific person, which is not applicable to other recognizers.

With the rapid development of semiconductor technology, the continuous increase in the scale of integrated circuits, and the continuous improvement of various development technology levels, English speech recognition technology has gradually become smaller after being combined with embedded systems based on DSP, FPGA, ASIC,

and other devices. With the development of industrialization and practicality, the application field is also getting bigger and bigger [3]. As a modern information technology with extensive social and economic benefits, English speech recognition has made certain achievements, but there are still a series of problems when facing practical use. The mature technology and reliable performance of English speech recognition systems are still available at home and abroad [4]. There is a lot of research space and market potential, and there is still a lot of room for improvement in terms of recognition accuracy, speed, robustness, and system miniaturization. In order to achieve an English speech recognition system with excellent performance, on the one hand, it is necessary to study the theory and algorithm of English speech recognition to solve and improve various problems in the recognition process. On the other hand, it is also necessary to consider simplifying the complexity of the system [5].

This paper compares the implementation schemes of the system and chooses to implement the English speech recognition system by way of SOPC. This article analyzes the requirements of the system, introduces the overall design of the system, and points out the components and software that a complete English speech recognition control system should include and the functional modules and software that the SOPC system should have. This paper divides the software and hardware that constitute the system and uses hardware to accelerate the algorithm in the most computationally expensive part of the software algorithm. At the same time, this article also introduces the selection of peripheral devices of the English speech recognition control system, the selection of processors, memories and buses that make up the SOPC system, and so on. This article selects the MicroBlaze soft processor core as the system processor, uses BRAM as the data storage memory, connects the data storage memory and other peripherals using the OPB bus, and connects the on-chip memory using the LMB bus. According to the large dependence of the training CHMM on the initial value and the large amount of calculation, an improved algorithm, Genetic Continuous Hidden Markov Model (GA_CHMM), is proposed. From the specific process of English speech signal preprocessing and endpoint detection, feature extraction, English speech recognition training, and recognition, the English speech recognition system based on the improved algorithm is explained. This paper tests the embedded English speech recognition system based on DSP and DHMM. For PC auxiliary software, the function of each module is mainly tested; after testing, each module can work normally. For the testing of embedded system software, the main indicators tested in this article are system recognition rate and real-time response. After testing, when the vocabulary of the system is 100, the recognition rate of the system can reach about 90%. After the optimization of fixed-point and CHMM algorithm, the average real-time response speed of the system has been improved.

2. Related Work

In order to ensure reproducible test results, avoid errors introduced by human factors, and improve test quality, the performance evaluation of embedded English speech recognition systems urgently needs to introduce automated test tools to replace heavy manual testing [6]. At present, the mainstream English speech recognition automated test method at home and abroad uses TTS technology to convert a text file containing test keywords into an English speech file and then uses a playback device to play the English speech file to perform English speech recognition. The test tool monitors the English speech recognition system in real time, obtains the recognition result, and records it to a file. Automatically they call the result statistics tool, compare the recognition results and the marked files, and determine whether the recognition results are correct or not. After the identification is completed, automatic summary statistics are performed, and a CSV file is output for easy viewing by testers [7].

The domestic research work on embedded software testing technology started from the field of defense electronics. Nanjing University, Beijing University of Aeronautics and Astronautics, and Aviation Industry Corporation have successively developed a number of testing tools and testing systems, whose main role is to cover embedded systems. After that, the employees of China State Shipbuilding Corporation conducted in-depth research on the GUI automated testing technology of embedded software and proposed a nonintrusive embedded GUI automated testing framework [8, 9]. Others proposed testing based on the characteristics of ship embedded software [10]. Scholars studied the UML-based embedded software test case generation technology [11]. Domestic embedded software testing tools generally target specific embedded operating systems, and most of them test for code coverage. It is difficult to have a clear understanding of the overall performance of embedded software.

The recognition performance test of the embedded English speech recognition system is carried out in the system test stage [12]. Based on the black box test method, the English speech recognition system is placed in the test environment (host machine environment or target machine environment) to simulate actual use and operation. Recognition tests are performed on batches of English speech data, and then the recognition performance indicators are counted based on the recognition results. At present, there are two main methods for testing embedded English speech recognition systems at home and abroad: playing test audio and speaking on the spot. Before the test, record the English speech data for the test in advance to prepare the noise data. In actual operation, a tester is required to operate the playback equipment, one tester to operate the English speech recognition system, and the test results greatly waste human resources and introduce errors in manual operation [13]. On-site oral calls are organized to organize multiple testers to read the test corpus of the English speech recognition system and test the effect. This method is too random and is not conducive to the recurrence of the test, and the way of pronunciation of different emotional infections will also be different, which may cause the instability of the test effect. Both test methods require live broadcast of English voice and noise signals. In the case of heavy test tasks and large

amounts of English voice data to be tested, it will undoubtedly cause a long test time and ultimately lead to the risk of a long product development cycle [14].

Using TTS technology to convert text into English voice can save the complexity of manual recording, but the converted English voice file is too "mechanized" and cannot simulate the tone and speed of human speech. In addition, the existing automated test methods still control the playback of English voice files by humans, and the time interval between each English voice is not flexible, which may lead to a waste of test time. The method is to compare the labeling data and the recognition results, but the labeling of English speech files is generally calculated by humans. Related scholars have proposed new ideas for applying vector quantization technology to English speech coding and recognition [15]. Hidden Markov model theory has become a research hotspot and an important theoretical basis for English speech processing technology [16]. Among them, the most representative is the Sphinx system, which is a nonspecific continuous English speech recognition system built using vector quantization and hidden Markov models. Three obstacles of English speech recognition were solved in the laboratory: nonspecific person, large vocabulary, and continuous English speech. In addition, artificial neural network technology has brought about new opportunities to the wide application of English speech recognition. Related scholars have designed an English speech recognition system based on the principles of the auditory nervous system [17–19]. Although the unique advantages of artificial neural network technology bring about many benefits to English speech recognition, the large amount of computation and long training time make its development relatively slow.

3. FPGA-Based Embedded Real-Time English Speech Recognition System Design

3.1. Hardware Platform Selection. The system design must meet the constraints of performance, cost, function, and so on. The overall design of the system is to divide the large and complex system into several modules according to actual needs and compare the advantages and disadvantages of the system composed of modules [20–22]. This article will compare the widely used embedded English speech recognition system and FPGA-based English speech recognition system in detail. Option one is a common English speech recognition system; the structure diagram is shown in Figure 1.

Solution one uses DSP digital signal processor or ARM processor as the central processing unit and CPLD as the coprocessor. But it is more common to use a DSP processor, because it can better reflect the advantages of digital signal processing. The second scheme uses the SOPC system composed of FPGA to form an English speech recognition system.

From the perspective of cost, the cost of a system with DSP/ARM as the processor is significantly higher than that of an English speech recognition system with FPGA as the core. A system with DSP/ARM not only requires a better processor but also requires a large amount of external equipment. The support of design increases the cost of system implementation. However, FPGA itself has a lot of logic resources, which can save the use of some memory and external logic and make the system miniaturized. From the performance point of view, the performance of the DSP/ARM system of the same price will be worse than that of the SOPC system. The use of functions makes it possible to improve system performance without increasing costs. Huawei, ZTE, and other companies use their own ASIC chips to reduce the cost by at least half compared with imported devices, and the performance obtained is indeed higher than the performance that can be obtained by using DSP platform. These advantages are not possessed by the system composed of DSP/ARM processor.

3.2. Design of English Speech Recognition System. The FPGA-based English speech recognition system is an embedded system that integrates software and hardware. It is an independent system that collects English speech, processes English speech data, and finally outputs control commands. In order to complete a larger-scale SOPC system, the design often adopts a top-down (Top-Down) hierarchical design idea. The hierarchical design idea is to divide a larger system into several subsystems, each subsystem is designed independently, and each subsystem is designed to be assembled into a complete system. The wiring between each functional submodule is as few as possible, the interface function is clear, and the scale of the functional module is required to be moderate. This design method greatly reduces the design complexity of system. A single subsystem can be tested separately. The problems that arise only need to modify the internal subsystems without affecting the functions of other subsystems.

System design should follow certain principles, use tools to divide modules, and determine what kind of functional modules the system should have and how to integrate these functional modules. The system design should solve the problem of the overall structure of the system in the hierarchical design of the system, rather than solving the problem of how to realize the partial functions. By organically integrating the divided system modules, the appropriate continuous method is used to maximize the system optimization. The system design uses a modular design method to grasp the framework of the entire system as a whole, so that the system meets the needs of the application. The system design follows the principle from top to bottom, decomposing each function one by one. The overall structure design of the English speech recognition system is shown in Figure 2.

The system architecture design consists of two parts: one is the architecture design of the entire English speech recognition system, and the other is the architecture design of the SOPC system. In Figure 2, the power supply is the energy source of the system, providing stable current and voltage for the system. The analog-to-digital conversion device converts the English voice signal into a digital signal and transmits it to the SOPC chip, so that the SOPC chip can process the English voice signal. In addition to software computing

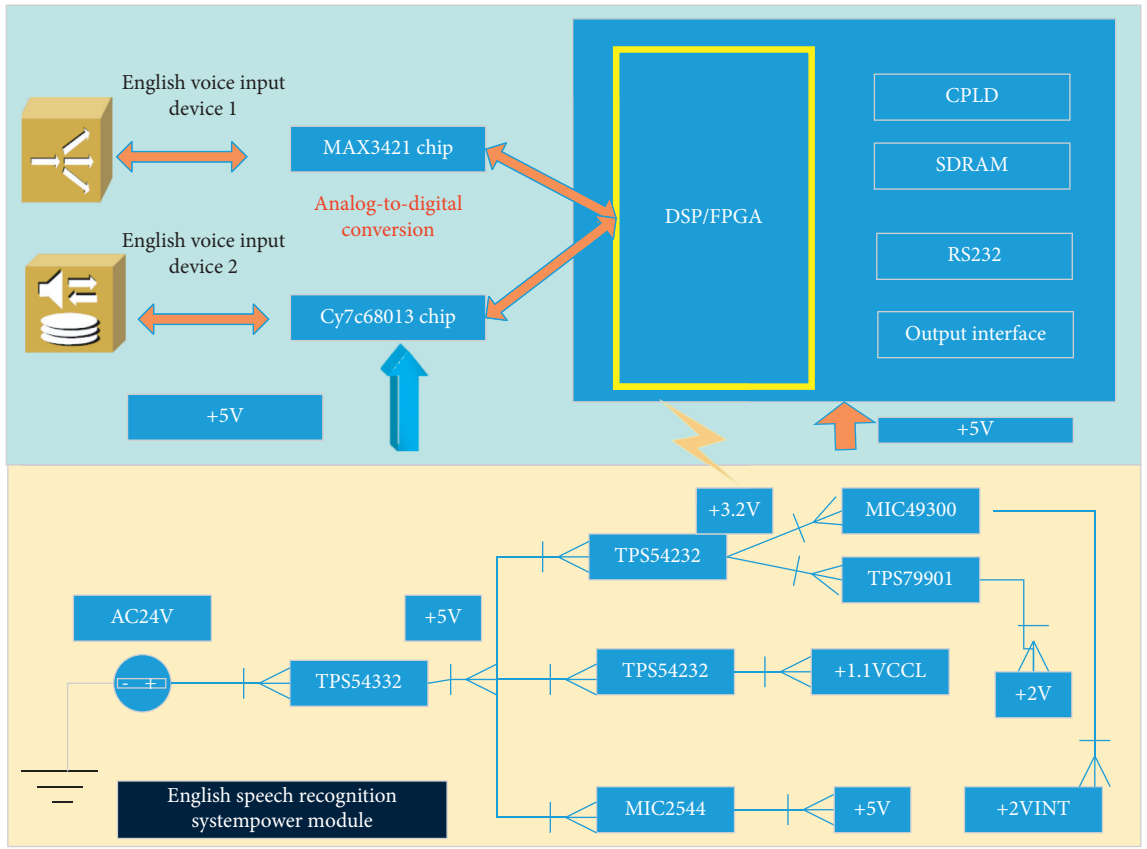


Figure 1: English speech recognition system scheme.

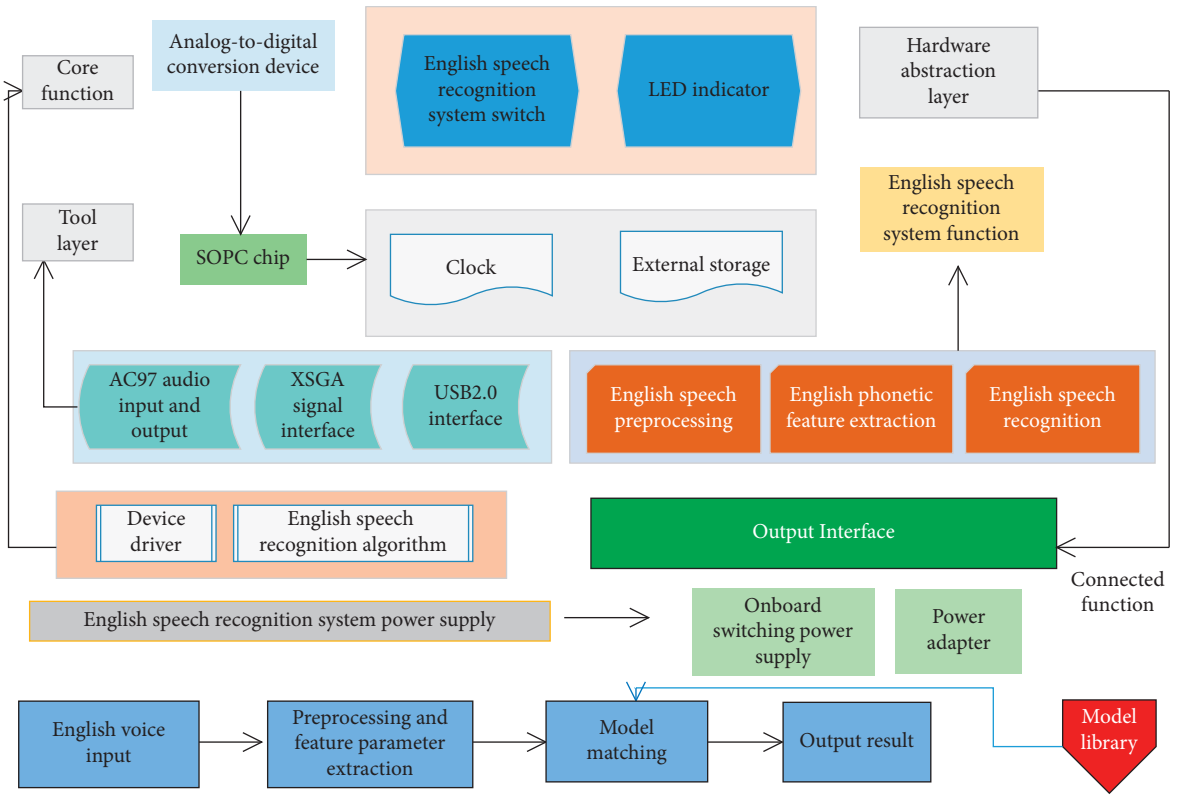


Figure 2: Structure diagram of English speech recognition system.

functions, SOPC chips also include hardware acceleration units. External storage is an indispensable part of a complete SOPC system, used to store models and data, programs, and so on. The output interface is the control signal interface output by the system, which provides the connection function between the system and other equipment.

3.3. SOPC System Design. The overall structure of the hardware design of the SOPC system is shown in Figure 3. It merges most of the devices that implement system functions to form a completed embedded system. The SOPC system accepts the input signal of the digital-to-analog conversion chip and outputs the control signal. The hardware part includes processor core, memory management IP core, memory, and hardware acceleration IP. At the same time, a complete SOPC system should also have an off-chip memory controller and so on.

As an integral part of the English speech recognition project, the part involved in the completion of this article includes program selection, overall system design, and system testing. The independently completed parts include software and hardware division, hardware configuration, English speech recognition algorithm recognition, software and hardware algorithm design optimization, hardware acceleration IP core, and its control logic design.

When designing an SOPC system, the software and hardware of the system are divided by requirements. The division of hardware, software, and hardware of the system is the key to whether the system can meet the needs of users. A 10% system-level design process has an 80% impact on the final cost and performance of the system design. Software is more flexible than hardware and easy to modify. However, because the software is executed programmatically, the speed is slower than the hardware. The use of hardware also has its disadvantages. The hardware method will increase the usage of SOPC system resources. When using FPGA to develop SOPC system, there may be the possibility that resources cannot meet the demand and require the use of higher-performance FPGA chips. If it is finally formed into an ASIC Chip, it may increase the area of the chip, thereby increasing the implementation cost of the system. Therefore, when realizing the goal of the system, the relationship between performance and cost must be considered comprehensively. After the hardware and software division of the system is completed, the designer then conducts the performance evaluation of the system. If the performance requirements are not met, the software and hardware must be redivided. To achieve the performance evaluation design, the hardware designer will consider the hardware physical realization of the system and go through the traditional IC design process, such as logic synthesis, layout planning, timing analysis, placement and routing, and physical verification. Software designers will consider which software operating environment to use, such as whether to use or not to use an operating system and which operating system to use.

There are multiple memory interfaces available on the XUPV2P platform, such as FLASH memory and DDR memory. The development tool provides IP cores for multiple memory controllers, and multiple memory controllers can be added to use these memories. OPB BRAM is faster than DDR memory, BRAM is faster, and the access of software and hardware to BRAM is much more efficient than that of DDR memory. BRAM has two ports, PORTA and PORTB; each port has an independent 32-bit address bus and 32-bit data bus, as well as read and write control lines. When implementing BRAM, four 8-bit RAMB16_s8_s8 devices are used to improve the parallelism of memory access. Only when each device is enabled, access is allowed. In SOPC design, external memory is often used to permanently save FPGA configuration and software data. This is because the contents of the FPGA will be lost after power failure. Therefore, we use external memory to save the configuration information and template data in the FPGA. At the same time, in order to speed up program execution, an on-chip memory of the processor should be set. The data used in the system is mainly divided into two types: one is model data, and the other is characteristic parameter data of samples. In order to facilitate the software and hardware to calculate the data and simplify the control operation, multiple BRAMs are designed in the system to save the model data transferred from the external memory, save the characteristic parameters of the samples, and save the results that need to be written back by the hardware acceleration calculation.

According to different application needs, MicroBlaze has a variety of buses available. The on-chip peripheral bus is a type of Core Connect bus. It is a low-speed bus that can be connected to buses of different widths and devices with different timings. In this article, the OPB bus is used to connect the hardware acceleration IP core control logic, because the OPB bus is mainly used for data transmission to external devices. The OPB bus supports multibit width data, which accepts input from the host when used for peripherals and performs specified operations, which meets the needs of hardware acceleration. OPB bus performance is shown in Table 1.

4. English Speech Recognition Algorithm Design Based on GA_CHMM

4.1. Parameter Selection of Genetic Algorithm. For genetic algorithms, choosing different control parameters will greatly affect the performance of the optimization. For the same encoding method and genetic operator, changes in parameters may cause greater performance changes. The control parameters of the genetic algorithm mainly include the length of the code string L, the population size N, the crossover probability P_c, the mutation probability P_m, and the termination algebra G.

① For length of encoding string L, regardless of whether the encoding method is real number encoding or binary encoding, according to the number of variables in the problem to be optimized and the encoding length L equal, each variable of the problem corresponds to each position in the encoding string. ② As regards group size N, the number of individuals in a group is called the group size. The size of

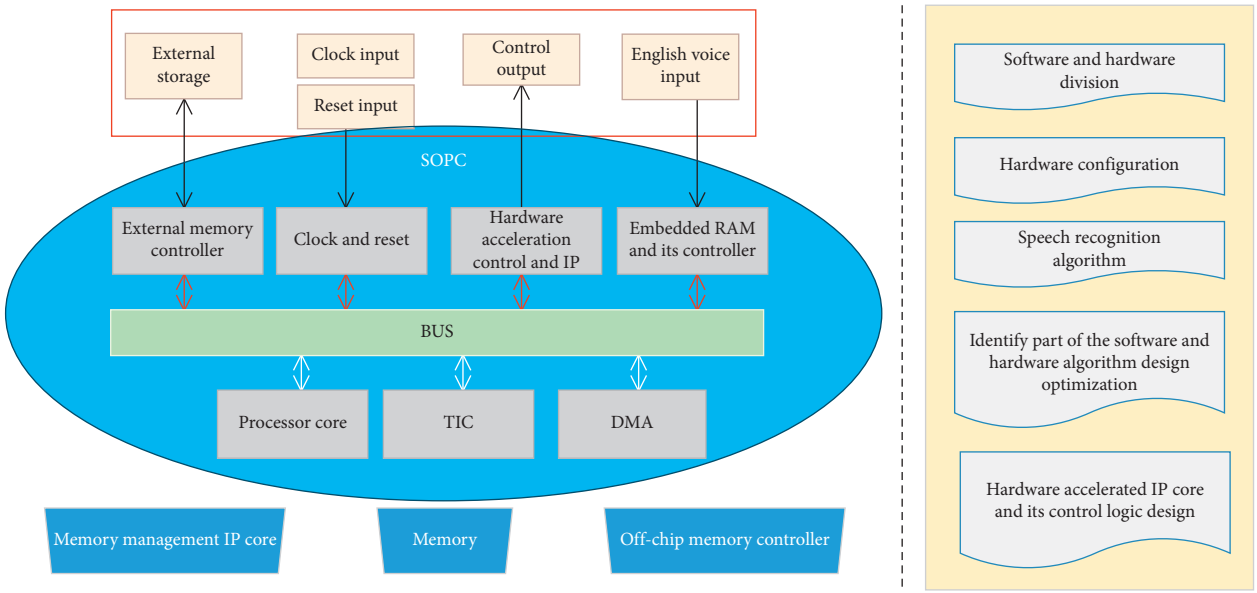


Figure 3: Overall scheme of SOPC system design.

Table 1: OPB bus performance.

Timing	Synchronize
Architecture	Multimaster/slave device
Interconnect	Bus independent read/write data line, does not support tristate
Connect	Multipath
Data line width	8-Bit, 16-bit, 32-bit
Data transfer protocol	Single read/write transmission, support burst transmission, word, byte, half word transmission
Address line width	32-bit

the group size N will affect the final result of optimization and the efficiency of the entire process. If the population size N is too small, the genetic algorithm cannot optimize the problem well; if the population size N is large, the genetic algorithm increases the probability of obtaining the local optimal solution, and the amount of calculation is relatively large, which will affect the optimization process of the entire process. So, depending on the actual situation, the value of N is also different, and generally the value is 20–100. ③ Regarding crossover probability P_c, the purpose of the crossover operation is to inherit the excellent genes in the parent to produce better individuals and get the best solution as possible in the iterative process. The crossover probability determines the search and optimization ability of the genetic algorithm. If P_c is small, the search and optimization ability of the genetic algorithm may fall into a relatively slow and sluggish state and cannot fully inherit the excellent genes; if P_c is large, the superior ability has been enhanced, but this may destroy the overall performance of the chromosome, so P_c is generally taken as 0.25–0.99. ④ For mutation probability P_m, the main purpose of mutation operation is to maintain the diversity of the population. When the mutation frequency P_m is small, although it can prevent the loss of important genes in the population, it reduces the possibility of population diversity. When the mutation frequency P_m is large, although the opportunity for diversity is increased, it may damage excellent individuals. Therefore, P_m is generally 0.0001~0.1. ⑤ For termination algebra G, it is a sign of the end of genetic algorithm. At this time, the best individual of the group is output as the optimal solution of the problem, and G is generally taken as 50~200.

4.2. Preprocessing of English Speech Recognition System Based on GA_CHMM. There are generally two methods for English voice collection: one is to use a hardware circuit system to collect English voice signals; the other is to directly use a multimedia sound card to collect English voice signals, that is, to use a computer sound card to collect English voice signals. The frequency of the English speech signal is within 40~4000 HZ. According to the Nyquist sampling theorem, the original English speech signal frequency is more than twice the sampling frequency for sampling. From the acquisition module, the sampling frequency is 11.025 kHz.

Through the preemphasis operation, the high-frequency part of the English speech signal is improved, and the power-frequency interference is filtered out to obtain a more pure and true English speech signal. Suppose that the English

speech sampling signal at time n is $x(n)$, and the signal obtained after preemphasis is $\hat{x}(n)$; that is,

$$\hat{x}(n) = nx(n) + 0.94x(n-1). \tag{1}$$

The transfer function of the preemphasis filter can be obtained as

$$H(z) = 1 + 0.94Z^{-1}. \tag{2}$$

After preemphasis, the frequency spectrum of the English speech signal is indeed improved in the high-frequency part, while filtering out the power-frequency interference. The English speech signal has short-term stability, so the English speech signal is divided into some equal time periods for analysis and processing, and this process can be achieved by using a movable finite-length window for weighting. Through the comparison of rectangular window, Hanning window, and Hamming window, this paper chooses Hamming window as the windowing function. $N = 256$ English speech samples are one frame. From the acquisition module, we know that the sampling frequency is 11.025 kHz, which can be calculated every time.

4.3. Optimization of HMM Parameters. According to the different description of the statistical characteristics of the observation sequence, there are two main types of HMM: discrete HMM (DHMM) and continuous HMM (CHMM). The biggest difference between the two is mainly in the method of calculating the probability B of the observation sequence. The parameter group B of DHMM is a probability matrix; that is, the probability $b_j(O_t)$ of an observation event produced by each state is satisfied:

$$\prod_{t=0}^{T-1} b_j(O_t) = -1. \tag{3}$$

The parameter group B of CHMM is that each state corresponds to an observation probability density function, which is a Gaussian probability density function:

$$b_j(O_t) = \prod_{m=0}^{M_j} N\left(U_{jm}, O_t, u_{jm}\right) \cdot C_{jm}. \tag{4}$$

CHMM does not need vector quantization. The mean value and variance are calculated from the feature vector after feature extraction, and the observation probability is calculated using the above formula. According to the needs of different practical problems, different HMM models can be selected.

Given the observation sequence O and the HMM model $\lambda = (A, B, \pi)$, the HMM parameter optimization problem is how to adjust the model $\lambda = (A, B, \pi)$ to maximize the probability of the observation sequence output $P(O, \lambda)$; here parameter reestimation is used to adjust the model parameters. The method of parameter reestimation is the Baum-Welch algorithm. The Baum-Welch algorithm continuously reevaluates the parameters through the reevaluation formula until convergence, the output probability $P(O, \lambda)$ is the largest, and then the parameter model obtained at this time $\lambda = (A, B, \pi)$ is the optimal HMM.

Given the observation sequence O and the model $\lambda = (A, B, \pi)$, the probability that the Markov chain is in the S_i state at time t and is in the S_j state at $t+1$ is as follows:

$$\xi_t(i, j) = P\left(S_i = q_t, S_j = q_{t+1}, O|\lambda - 1\right). \tag{5}$$

Then we launch

$$\xi_t(i, j) = a_{ij} b_j(O_t) \frac{\alpha_t(i)\beta_{t+1}(j)}{P\left(S_i, S_j, O|\lambda - 1\right)}. \tag{6}$$

The probability that the Markov chain is in S_i at time t is

$$\xi_t(i) = \prod_{j=0}^{N-1} \xi_t(i, j) = \frac{\alpha_{t+1}(i)\beta_t(i+1)}{P(O|\lambda - 1)}. \tag{7}$$

4.4. The Training Process of English Speech Recognition. The English speech signal is a time series. According to the characteristics of the English speech signal, each word can be represented by a CHMM model parameter. The system adopts the HMM model with $N = 5$ states from left to right without spanning. After the feature parameters of the English speech signal are extracted, a word feature vector is used as the input observation sequence of the CHMM model.

The CHMM training algorithm (Baum-Welch algorithm) is to iteratively calculate the observation sequence of the English speech signal through an estimation formula to obtain a new parameter model. The new parameter model will be better than the old parameter model. Through repeated iterations until the convergence condition is reached, the best CHMM model is obtained. The best model at this time is most likely to be a local optimal solution rather than a global optimal solution. In order to obtain the global best model as much as possible and obtain a better recognition effect, genetic algorithm is introduced in the process of CHMM training; that is, a new training algorithm GA_CHMM is obtained. GA_CHMM algorithm is realized from the following aspects.

4.4.1. Coding Scheme. In the process of applying genetic algorithm to CHMM training, the parameters that need to be optimized in the CHMM model are first arranged to form a chromosome. This paper uses the CHMM model with five states jumping from left to right. The parameters of the CHMM model mainly include the initial state distribution probability π, the state transition matrix probability A, and the probability density function B of each state corresponding to the observation sequence. There are 5 parameters in the initial state distribution matrix π. There are a total of $5 \times 5 = 25$ parameters in the state transition matrix A, and the mixing coefficient matrix C in the probability density function B has $5 \times 5 = 25$ parameters. In this paper, 24 order MFCC coefficients are used. CHMM does not need vector quantization. The feature vector after feature extraction is used to obtain the mean and covariance through the

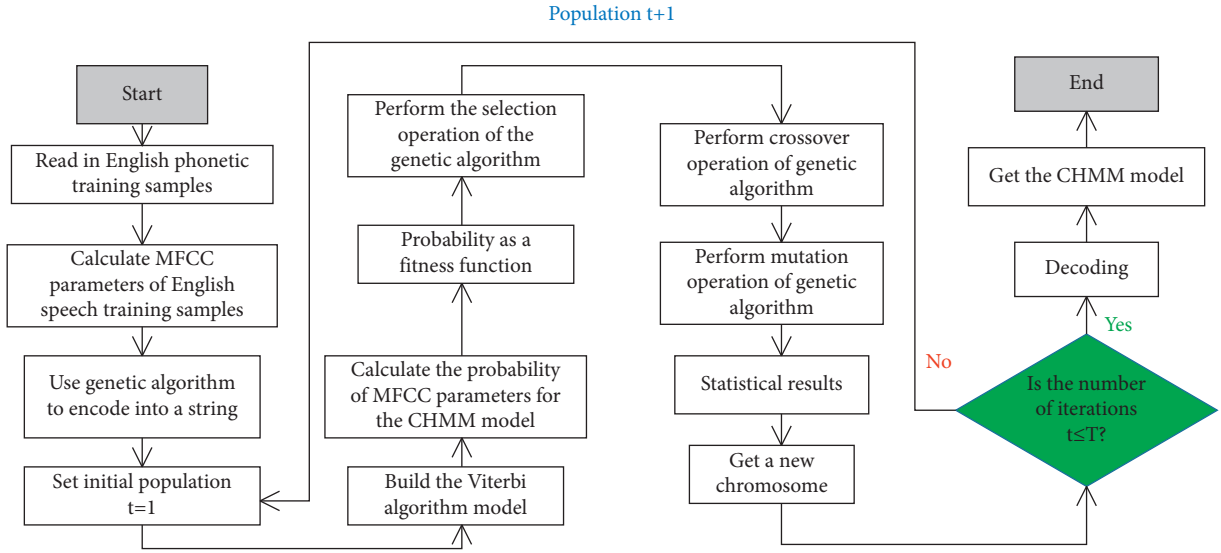


Figure 4: Training flowchart.

probability density function B. Mean μ and Covariance U are $5 \times 5 \times (24 + 24 \times 24) = 15000$ parameters which are combined into a string in rows to form the back part of the chromosome, so that the front part and the back part together form a chromosome. The sum of each row vector of the mixing coefficient C matrix is 1.0. After the genetic operation, parameters A and C must be normalized.

4.4.2. Fitness Function. The higher the likelihood of the training data to the model, the better. Here, the fitness function of the individual chromosome is expressed by the log-likelihood probability of the observed sequence of feature parameters of the English speech signal; namely,

$$f(\lambda) = \ln[P(O(k)|\lambda - 1)]. \quad (8)$$

In the above formula, $O(k)$ represents the k-th observation sequence used to train the model.

4.4.3. Population Initialization. The population initialization produces 100 chromosomes; that is, the population size is 100. Based on the fitness function, the fitness value is compared according to the fitness function size of each chromosome. 40 excellent chromosomes were selected from the population directly as part of the next-generation chromosomes. In addition, another part of 60 chromosomes is generated through crossover and mutation operations, which together form a new generation of chromosomes.

After real-number encoding of the chromosomes, the length of the chromosome is $L = 15055$. Through population initialization, 100 chromosomes are generated. According to the fitness value, 40 excellent chromosomes are selected directly as the next-generation chromosomes. The chromosomes of the next-generation population are better than the previous-generation chromosomes, so, after repeated iterations until $G = 60$, the generation is terminated, and the corresponding model is the CHMM model. The training process of GA-CHMM is shown in Figure 4.

5. System Test and Result Analysis

5.1. System Recognition Rate Test

5.1.1. Test and Analysis of the System's Different Vocabulary Recognition Rate. Since the system can set the size of the selected isolated vocabulary, different vocabularies of different sizes are selected for the recognition rate of the system during the test. The size of the vocabulary selected for the test is 5~200. When the system tests the recognition rate on the embedded system, it uses the PC-assisted software to complete the collection of English speech samples, the training of the template, and the update of the template data.

Before the test, first collect the English speech samples of isolated words. When collecting samples, 10 people are selected, and 4 English speech samples are collected for each isolated word; in this way, there are a total of 40 samples for each isolated word to participate in the training. After the English speech samples of all isolated words are collected, the English speech template is trained according to the set vocabulary size and then downloaded to the embedded system through the USB cable to test the recognition rate of the system. In each test, the tester spoke to each isolated vocabulary 10 times and recorded the system's recognition of the isolated vocabulary. We divide the number of successful recognitions by the total number of speeches to get the recognition rate of the system test. The experimental test results of the system's recognition rate of different vocabularies are shown in Figure 5.

The average recognition rate in Figure 5 is the average of the recognition rates of three experimental tests for each vocabulary. It can be seen from the results in Figure 5 that the recognition rate of the system changes tortuously as the vocabulary increases. When the vocabulary reaches 200, the recognition rate of GA_CHMM algorithm is about 88%.

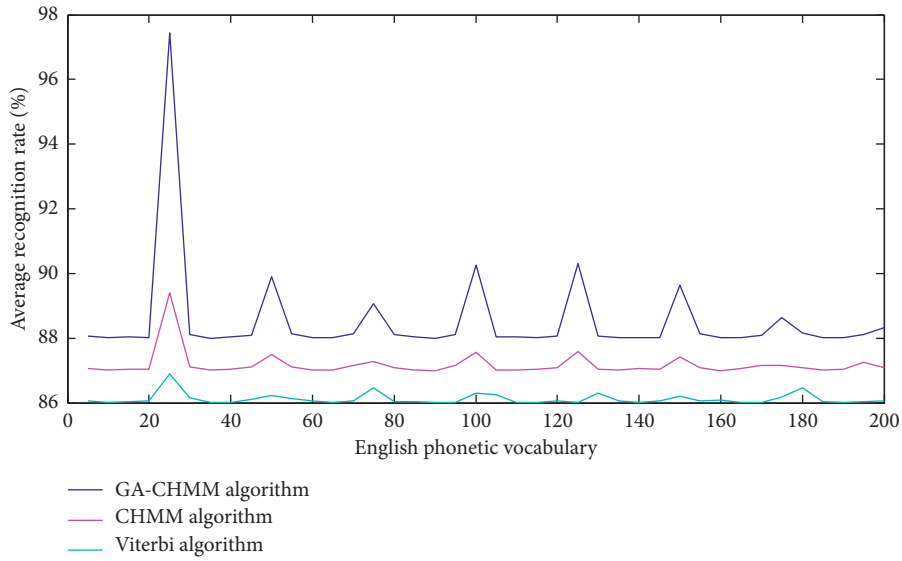


Figure 5: Recognition rate test results of different vocabularies.

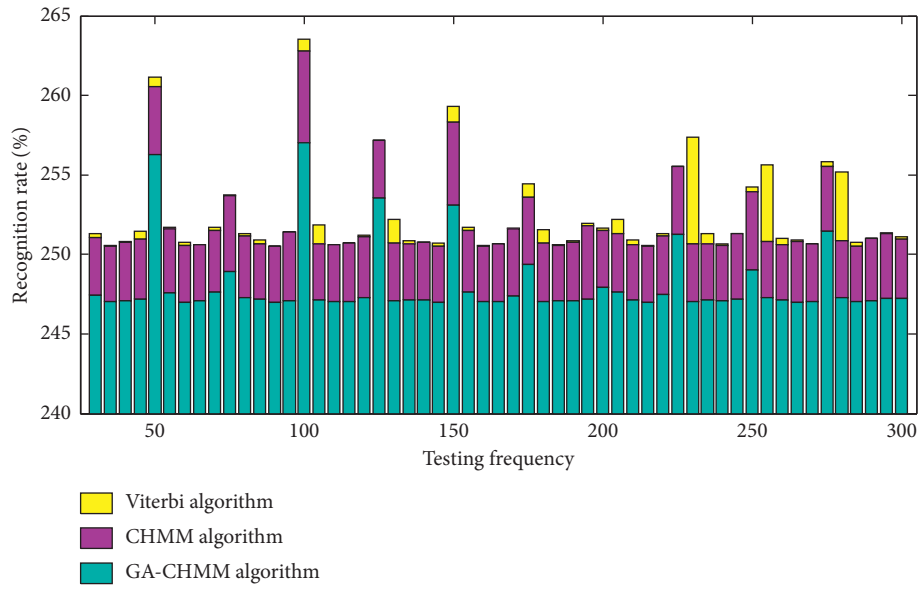


Figure 6: The test result of the recognition rate of the system nonspecific person.

5.1.2. Test and Analysis of the System's Recognition Rate of Unspecified Persons. The system selects 200 vocabularies, and 10 people are selected for the test. First select 8 people from 10 to train the English phonetic template, and each person samples 4 for each isolated vocabulary. In this way, there are 32 English phonetic samples for each isolated vocabulary. The other two people were used as the system unspecified person test, each isolated word was said 3 times, each person said the words 300 times in total, and the system's recognition rate was counted; the experimental results are shown in Figure 6.

This system provides the learning function of isolated words in English speech recognition. For vocabulary whose recognition rate is not high in the recognition process, the learning function is used to learn this vocabulary. After the learning is completed, we test the recognition rate of the system. The experimental test results are shown in Figure 7.

The experimental results in Figure 7 show that, after the learning function, the recognition rate of the system has been greatly improved. After the learning of isolated words, the recognition rate has reached about 90%. The system can achieve the learning function of isolated words through the learning function.

5.2. Real-Time Testing of the System. The vocabulary size of isolated words used in the experiment is 100. The time taken by the system from the completion of the collection of English speech to the recognition of the result is recorded as the system's response time.

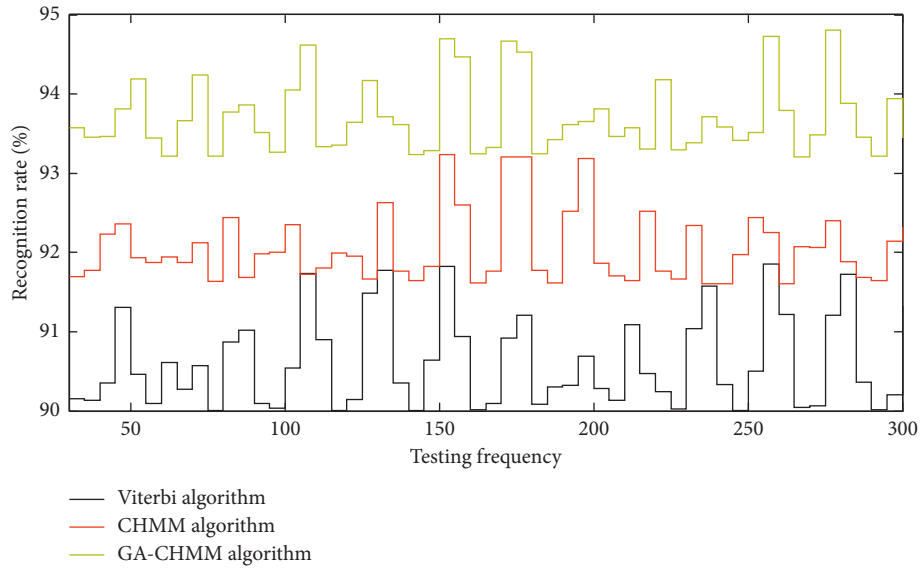


Figure 7: Recognition rate test results after using the learning function.

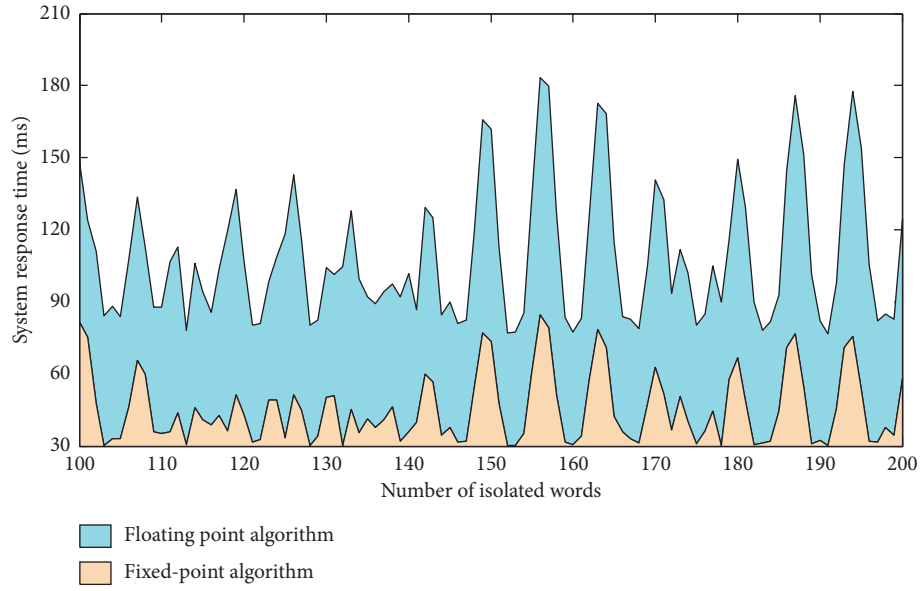


Figure 8: The system's response time using floating-point arithmetic and fixed-point arithmetic.

5.2.1. The Response Time of the Test System in the Case of Fixed-point and Floating-Point Arithmetic. We select 100 to 200 isolated words as test objects and test the system's response time when each word is recognized on the embedded system software, using floating-point arithmetic and fixed-point arithmetic, respectively. The test result is shown in Figure 8.

It can be seen from the results shown in Figure 8 that the system has a maximum response time of 180 ms when using floating-point arithmetic. After adopting fixed-point algorithm, the real-time performance of the system has been greatly improved, and the maximum response time of the system is 85 milliseconds.

5.2.2. The Response Time of the Test System Using Different Recognition Algorithms. We set the vocabulary of the system to 100 and perform English speech recognition for each vocabulary and record the real-time response time of GA_CHMM algorithm, CHMM algorithm, and Viterbi algorithm, respectively. The test results are shown in Figure 9.

In the test results shown in Figure 9, after the system adopts the optimized CHMM algorithm, the average response time of the system is the lowest, while the average response time of the system using the unoptimized CHMM algorithm is higher; the response time of the Viterbi algorithm is the highest.

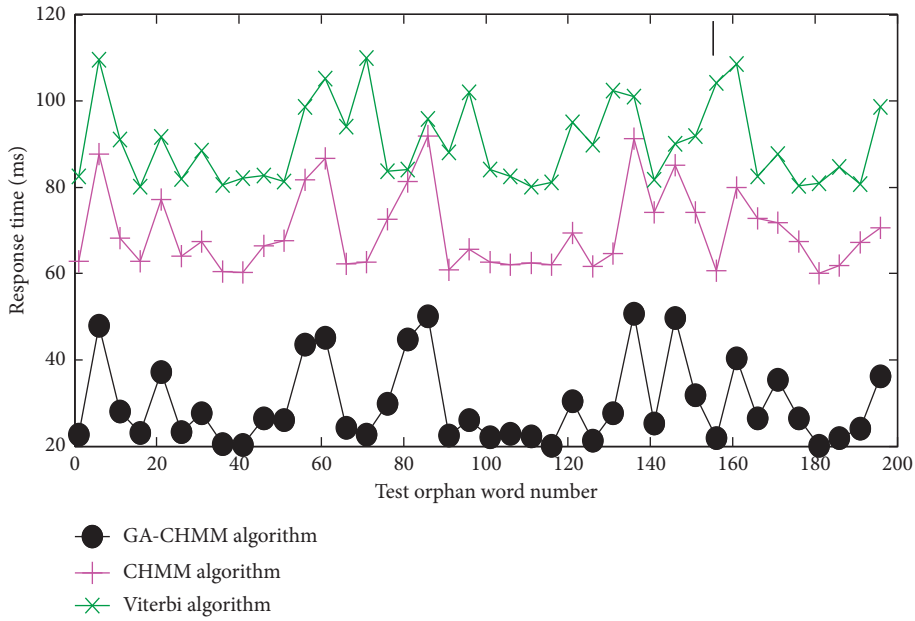


Figure 9: Response time of system algorithm.

6. Conclusion

This text has carried on the overall design of the English speech recognition system based on FPGA. We choose to use SOPC this way to realize the system and choose the software module through the comparison of algorithms. Equipped with the English speech recognition control system and the hardware part of the SOPC system, the software and hardware modules of the SOPC system are divided. The processor and memory of the hardware acceleration unit are configured in detail, and a memory implementation method is designed according to the needs of the system. The design of English speech recognition control system and hardware acceleration unit IP core is realized. This article analyzes the data and gives the scope of the data and draws a way of data representation. Combining modules from top to bottom describes the design of hardware IP. The genetic algorithm has the characteristics of superior global search capability and parallel computing. It improves the traditional CHMM algorithm and uses the genetic algorithm to directly train the CHMM model. This process mainly includes coding method, fitness function design, population initialization, selection operation, crossover operation, mutation operation, and termination strategy. The English speech recognition system based on the improved algorithm (GA_CHMM) is studied, which mainly includes the establishment of English speech template, preprocessing, endpoint detection, feature extraction, and training process. The English speech recognition system was simulated by MATLAB software, and a better recognition effect was obtained. This paper studies the real-time index of embedded English speech recognition system and the system optimization method of recognition rate. Through the analysis of the system hardware, technical methods such as fixed-point embedded system algorithm and storage space optimization are adopted to ensure the real-time requirements of the system. At the same time, this article focuses on the optimization of the CHMM algorithm in the recognition phase and proposes an optimized CHMM algorithm based on the irreversible characteristics of the English speech signal, and the real-time performance of the system is once again improved.

Data Availability

All the data are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

References

[1] B. Kurniadhani, S. Hadiyoso, and S. Aulia, "FPGA-based implementation of speech recognition for robocar control using MFCC," *Telkomnika*, vol. 17, no. 4, pp. 1914–1922, 2019.

[2] A. A. Gilan, M. Emad, and B. Alizadeh, "FPGA-based implementation of a real-time object recognition system using convolutional neural network," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 4, pp. 755–759, 2019.

[3] T. Posewsky and D. Ziener, "Throughput optimizations for FPGA-based deep neural network inference," *Microprocessors and Microsystems*, vol. 60, pp. 151–161, 2018.

[4] M. A. A. de Sousa, R. Pires, and E. Del-Moral-Hernandez, "SOM processor: a high throughput FPGA-based architecture for implementing Self-Organizing Maps and its application to video processing," *Neural Networks*, vol. 125, pp. 349–362, 2020.

[5] M. Ling, M. Javad Esfahani, H. Akbari, and F. Amin, "Effects of residence time and heating rate on gasification of petroleum residue," *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1837–1840, 2016.

[6] H. L. Ma and S. B. Tsai, "Design of research on performance of a new iridium coordination compound for the detection of

Hg2+," *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, p. 1232, 2017.

[7] L. Y. Mo, W. H. Z. Sun, S. Jiang et al., "Removal of colloidal precipitation plugging with high-power ultrasound," *Ultrasonics Sonochemistry*, vol. 69, p. 105259, 2020.

[8] X. Yang, Q. Zhou, J. Wang et al., "FPGA-based approximate calculation system of General Vector Machine," *Microelectronics Journal*, vol. 86, pp. 87–96, 2019.

[9] C. Gu, N. Hanley, and M. O'neill, "Improved reliability of FPGA-based PUF identification generator design," *ACM Transactions on Reconfigurable Technology and Systems*, vol. 10, no. 3, pp. 1–23, 2017.

[10] L. Xia, L. Diao, Z. Jiang et al., "Pai-FCNN: Fpga based inference system for complex CNN models,"vol. 2160, pp. 107–114, in *Proceedings of the 2019 IEEE 30th International Conference on Application-Specific Systems, Architectures and Processors (ASAP)*, vol. 2160, pp. 107–114, IEEE, New York, NY, USA, July 2019.

[11] S. Afifi, H. GholamHosseini, and R. Sinha, "A system on chip for melanoma detection using FPGA-based SVM classifier," *Microprocessors and Microsystems*, vol. 65, pp. 57–68, 2019.

[12] M. Sahani and P. K. Dash, "FPGA-based online power quality disturbances monitoring using reduced-sample HHT and class-specific weighted RVFLN," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4614–4623, 2019.

[13] D. Xu and H. Ma, "Degradation of rhodamine B in water by ultrasound-assisted TiO_2 photocatalysis," *Journal of Cleaner Production*, vol. 313, Article ID 127758, 2021.

[14] D. Gao, Y. Liu, Z. Guo et al., "A study on optimization of CBM water drainage by well-test deconvolution in the early development stage," *Water*, vol. 10, no. 7, 2018.

[15] H. Y. Kim, L. Xu, W. Shi et al., "A secure and flexible FPGA-based blockchain system for the IIoT," *Computer*, vol. 54, no. 2, pp. 50–59, 2021.

[16] K. K. Guner, T. O. Gulum, and B. Erkmen, "FPGA-based wigner–hough transform system for detection and parameter extraction of LPI radar LFMCW signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–15, 2021.

[17] J. W. Chang, K. W. Kang, and S. J. Kang, "An energy-efficient FPGA-based deconvolutional neural networks accelerator for single image super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 281–295, 2018.

[18] O. Oballe-Peinado, J. A. Hidalgo-Lopez, J. Castellanos-Ramos et al., "FPGA-based tactile sensor suite electronics for real-time embedded processing," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 12, pp. 9657–9665, 2017.

[19] C. Xu, Z. Peng, X. Hu et al., "FPGA-based low-visibility enhancement accelerator for video sequence by adaptive histogram equalization with dynamic clip-threshold," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 11, pp. 3954–3964, 2020.

[20] T. Grubljesic, P. S. Coelho, and J. Jaklic, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2019.

[21] L. X. Z. Zhang, M. Mouritsen, and J. R. Miller, "Role of perceived value in acceptance of bring your own device policy," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 65–82, 2019.

[22] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer gamification: the consensus, the best practice and the grey areas," *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9989731, 8 pages
https://doi.org/10.1155/2021/9989731

Research Article

Research on Ship Meteorological Route Based on A-Star Algorithm

Ge Chen,[1] **Tao Wu**,[1,2] **and Zheng Zhou**[3]

[1]*Chengdu University of Information Technology, School of Computer Science, Chengdu 610225, China*
[2]*Sichuan Numerical Weather Computing Engine Research Center, Chengdu 610225, China*
[3]*Ninecosmos Science and Technology Ltd., Wuxi 21400, China*

Correspondence should be addressed to Tao Wu; wut@cuit.edu.cn

Received 28 March 2021; Accepted 30 April 2021; Published 18 May 2021

Academic Editor: Xianyong Li

Ship meteorological navigation is based on hydrometeorological data of a certain time scale, considering the ship's motion characteristics and its own characteristics. First, we provide the best route for the ship and then use real-time local weather information to correct the route during the ship's navigation. It can also be expressed as follows: it is based on the hydrological and meteorological conditions of the ship during its voyage and the seakeeping characteristics of the ship itself, and the route planning method is used to select the best route for the ship. The best route is a balance between economy and safety, that is, based on ensuring the safety of ship navigation, the route that meets the shortest navigation time, the least fuel consumption, or the least navigation risk is obtained. Weather navigation includes the optimization of the initial route before sailing and the correction of the route after sailing. As there may be errors in hydrometeorological forecasts, especially in the accuracy and real-time performance of medium and long-term forecasts, the optimal initial route may not achieve the best results. Therefore, after the ship sails, it is necessary to adjust and correct the preferred initial route based on the meteorological information detected by the sensors or the continuously updated hydrometeorological forecast data to ensure the best effect of meteorological navigation. This paper proposes a weather route planning method based on the improved A-star algorithm. The convex shape of the concave obstacle and the expansion of the obstacle are carried out; according to the position of the target point relative to the starting point, the search direction of the A-star algorithm at each node is restricted, and an improved A-star algorithm is proposed. The simulation of global weather route planning shows that the improved A-star algorithm can not only find the optimal path but also effectively reduce the number of nodes that the algorithm needs to search during operation. Compared with the classic algorithm, the improved algorithm reduces the number of node searches by 29.25%.

1. Basic Theory and Methods of Meteorological Navigation

1.1. Overview of Meteorological Navigation. Ship meteorological navigation is based on the hydrometeorological data of a certain time scale, taking into account the ship's motion characteristics and its own characteristics. First, we provide the best route for the ship and then the navigation technology of correcting the route by using real-time local meteorological information in the course of ship navigation [1]. Meteorological navigation includes the selection of the initial route before sailing and the correction of the route after sailing. Due to the error of hydrometeorological forecast, especially the deviation of accuracy and real-time of medium and long-term forecast, the optimal initial route may not achieve the best effect. Therefore, when the ship is sailing, it is necessary to adjust and correct the selected initial route according to the meteorological information detected by the sensor or the continuously updated hydrometeorological forecast data, to ensure the best effect of meteorological navigation.

1.2. Characteristics of Meteorological Navigation.

(1) All hydrological and meteorological conditions are considered in the route planning before and after the voyage. Therefore, the meteorological navigation can prevent the ship from navigating in areas with heavy wind and waves to ensure the safety of the ship [2].

(2) Navigation safety is the prerequisite for route planning, and weather navigation can plan different routes for ships with different sailing goals [2].

1.3. Presentation of Environmental Models. Cognitive map is a spatial model, which expresses the interconnection between objective things and their geographical distribution. Environmental modeling is the embodiment of the characteristics of the surrounding environment, and it is a cognitive map based on perceptual knowledge [3]. Most of the environmental modeling in route planning is to transform environmental information into maps, that is, to establish a network model. Common methods include grid-based map method [4], geometric feature map [5] method, and topological map method [6]. Montes [7] used a square grid map composed of longitude and latitude to divide the Western Pacific and established an *OSTR* (*optimum track ship routing*) model for the US Navy fleet. This paper also uses square grid graphs to represent the environment model. Since the research focus of this article is on the algorithm of route planning and does not consider the specific movement characteristics of the ship, this article simplifies the ship in a two-dimensional space and approximates it to a point to represent.

As shown in Figure 1, the ship's current position is $M(\lambda_i, \phi_i)$, Use the longitude lines from λ_1 to λ_n and the latitude lines from ϕ_1 to ϕ_n to divide the sea area containing the starting point $S(\lambda_1, \phi_1)$ and the target point $E(\lambda_n, \phi_n)$. The accuracy of the grid can be $0.5^\circ \times 0.5^\circ$, $1^\circ \times 1^\circ$, $2^\circ \times 2^\circ$, etc. A grid diagram is a directed network that can be defined as follows:

$$G = (N, A), \tag{1}$$

where N is the set of nodes and A is the set of directed arcs [8]. Each intersection of the warp and weft is a node. Suppose the accuracy of the square grid is $s^\circ \times s^\circ$, number the nodes in the order from bottom to top and from left to right, for example, the starting point S is numbered 1, and the current point of ship M is numbered $(((\phi_n - \phi_1)/s) + 1) \times (\lambda_i - \lambda_1/s) + (((\phi_1 - \phi_1)/s) + 1)$. As equation (2) and (3), if the node number is N_0, then the latitude and longitude of the node has the following relationship with its number:

$$\lambda = \lambda_1 + \frac{N_0 - 1}{t} \times s, \tag{2}$$

$$\phi = \phi_1 + \left[(N_0 - 1) \bmod t\right] \times s. \tag{3}$$

Inside equation (3), $t = ((\phi_N - \phi_1)/s) + 1$.

The simulations in this article use a square grid graph with a moving direction, that is, each node has 8 out-degrees and 8 in-degrees [9], except for nodes on the border and corners, as shown in Figure 1. T_1, T_2, T_3, T_4, T_5, T_6, T_7, and T_8 are the current node M where the ship is in order, starting from true north and clockwise at every 45° out of 8 directions and satisfying $T_1 = 0^\circ$, $T_2 = 45^\circ$, $T_3 = 90^\circ$, $T_4 = 135^\circ$, $T_5 = 180^\circ$, $T_6 = 225^\circ$, $T_7 = 270^\circ$, and $T_8 = 315^\circ$. Due to the limitation of the boundary of the navigation area, the corner nodes have only 3 out-degrees and 3 in-degrees, and the nodes on the boundary have only 5 out-degrees and 5 in-degrees [10]. For

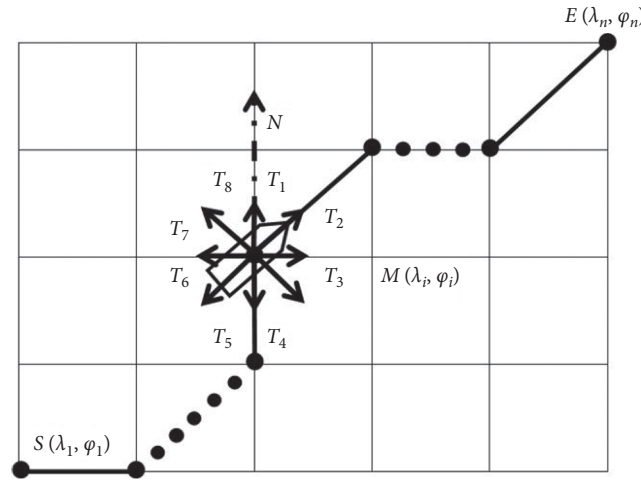


Figure 1: Grid-based map.

a node numbered N_0 with 8 out-degrees, the numbers of adjacent nodes in the 8 out-degree directions are $N_0 + 1$, $N_0 + t + 1$, $N_0 + t$, $N_0 + t - 1$, $N_0 - 1$, $N_0 - t - 1$, and $N_0 - t + 1$.

Inside, $t = ((\phi_N - \phi_1)/s) + 1$.

2. A-Star Algorithm Theoretical Basis

2.1. Dijkstra's Algorithm and Best First Search Algorithm. Dijkstra's algorithm visits the nodes in the graph from the starting point [11]. It constantly checks the nodes adjacent to the current node in the set of all nodes that have not found the shortest path and adds the node with the smallest cost value from the set of all nodes that have not found the shortest path to all the nodes that have found the shortest path. In the collection, it starts from the starting point and expands layer by layer until it finds the target point. If all edges in the graph have nonnegative cost values, Dijkstra's algorithm guarantees to find the shortest path from the starting point to the target point [12]. As shown in Figure 2, the green grid is the location of the starting node, the red grid is the location of the target node, the light blue area represents the area searched by Dijkstra algorithm, the light green grid represents the boundary of the search node area, and the light-yellow dashed line is the shortest path found by Dijkstra algorithm.

The steps of the best first search (BFS) algorithm are like the Dijkstra algorithm, and it can estimate the moving cost of the node to the target point. It favors nodes close to the target point rather than nodes close to the starting point. It runs fast, but there is no guarantee that the path found is the shortest [13]. The heuristic function used by the best first search algorithm can make it quickly approach the target point, and it runs faster than the Dijkstra algorithm under the same conditions. As shown in Figure 3, the green grid is the location of the starting node, the red grid is the location of the target node, the light blue area represents the area searched by best first search algorithm, the light green grid represents the boundary of the search node area, and the light-yellow dashed line is the shortest path found by best first search algorithm.

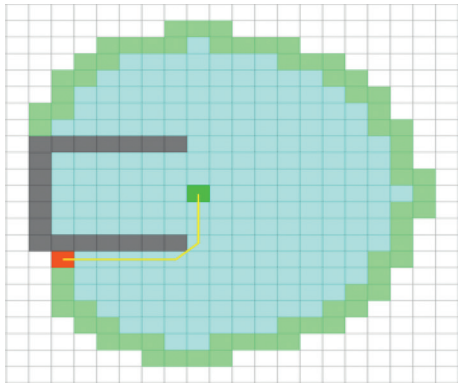


Figure 2: Route planning simulation of Dijkstra algorithm.

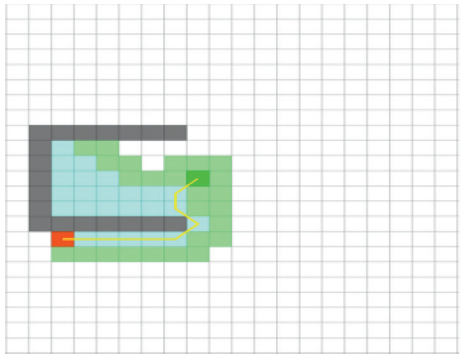


Figure 3: Route planning simulation of BFS algorithm.

By comparing Figures (2) and (3), we can find (1) Dijkstra algorithm can indeed find a shortest path; (2) the path found by the best first search algorithm is obviously not the shortest path; (3) the grid area scanned by the best first search algorithm is significantly less, so it runs faster than Dijkstra's algorithm.

2.2. A-Star Algorithm. The A-star algorithm concentrates the advantages of both the Dijkstra algorithm and the best first search algorithm. It is an algorithm that combines the advantages of the conventional algorithm (such as the Dijkstra algorithm) and the heuristic algorithm (such as the best first search algorithm) [14]. A-star algorithm is one of the most popular algorithms in path planning. It is quite flexible and can be used in various scenarios. Like the Dijkstra algorithm, the A-star algorithm can be used to find the shortest path [15]. Like the best first search algorithm, the A-star algorithm can use heuristic functions to guide the approach to the target point. In the simplest case, where there are no obstacles, the A-star algorithm can be as fast as the best first search algorithm. In the presence of concave obstacles as shown in Figures (2)–(4), the A-star algorithm can find the same good results as the Dijkstra algorithm. In Figure 4, the green grid is the location of the starting node, the red grid is the location of the target node, the light blue area represents the area searched by A-star algorithm, the light green grid represents the boundary of the search node area, and the light-yellow dashed line is the shortest path found by the A-star algorithm.

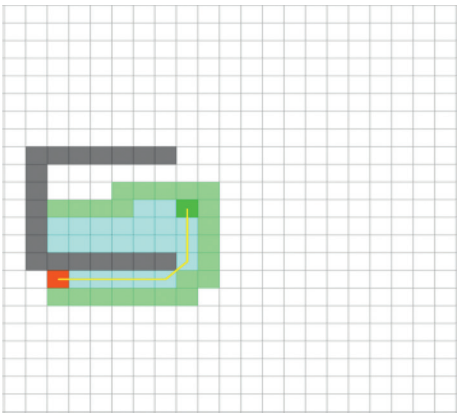


Figure 4: Route planning simulation of A-star algorithm.

The A-star algorithm needs to use the cost function as follows:

$$F(\mathrm{n}) = G(\mathrm{n}) + H(\mathrm{n}), \tag{4}$$

where $F(n)$ is the estimated value of moving from the starting point to the nth node, $G(\mathrm{n})$ is the cost value of moving from the starting point to the nth node along the generated path, and $H(n)$ is the heuristic function, used to estimate the cost value of moving from the nth node to the target point.

The A-star algorithm also needs to use the open table and closed table: the open table is used to store the nodes that have bee.

The execution process of the A-star algorithm is as follows:

(1) Add the starting point to the open table

(2) Repeat the following tasks:

  (1) Find the node with the lowest F value in the open table and set the current node as the node with the lowest F value.
  (2) Remove the current node from the open table and put it into the closed table.
  (3) Judge all nodes adjacent to the current node in order. (c1) If the node is already in the closed table, then the node is not considered. (c2) If the node is not in the open table, add the node to the open table and use the current node as the node the parent node of the node and record the value of F, G, and H of the node. (c3) If the node is already in the open table, the value of G that reaches the node after passing through the current node is less than the value of G that reaches the node without passing through the current node; the current node is taken as the parent node of the node, and the F, G, and H values of the node are recalculated. Otherwise, the F, G, and H values of the node remain unchanged.
  (4) Stop searching: if the target point is already added to the closed table, the path is saved; if the target point is not added to the closed table and

the open table is empty, the path does not exist generated but not visited, and the closed table is used to store the nodes that have been visited.

2.3. The Influence of Heuristic Function H (n) on A-Star Algorithm. Whether the A-star algorithm can guarantee to find the optimal solution, the key lies in the selection of the heuristic function $H(n)$. For the evaluation function: $F(n) = G(n) + H(n)$, the selection of the heuristic function $H(n)$ has the following situations:

(1) In a limit case, if $H(n) = 0$, then $F(n) = G(\mathrm{n})$, that is, only $G(n)$ works. At this time, the A-star algorithm is equivalent to the Dijkstra algorithm, and the path found is the shortest path.

(2) If $H(n)$ is always less than the cost of moving from node n to the target point, then the path found by the A-star algorithm is also the shortest path. Moreover, the smaller the value of $H(n)$, the more nodes the A-star algorithm expands, and the slower the execution speed of the A-star algorithm.

(3) If $H(n)$ is exactly equal to the cost of moving from node n to the target point, then the A-star algorithm will only move along the optimal path without expanding other irrelevant nodes. At this time, the A-star algorithm will run very fast. Although this cannot happen in all scenarios, it can happen in some special scenarios.

(4) If $H(n)$ is sometimes more expensive than moving from node n to the target point, the A-star algorithm cannot guarantee to find a shortest path, but it runs faster.

(5) In another extreme case, if $H(n)$ is much larger than $G(n)$, that is, $F(n) \approx H(n)$, only $H(n)$ will work. At this time, the A-star algorithm will approximate for the best first search algorithm, and this situation cannot guarantee to find a shortest path, but the algorithm runs fast.

Ideally, we want to get the shortest path the fastest. If the value of $H(n)$ is too small, then we can get the shortest path, but the running speed of the algorithm will become very slow. If the value of $H(n)$ is too large, then we must abandon the shortest path, but the algorithm will run faster. We need to make a trade-off between the speed of the algorithm and the accuracy of the results. In some application scenarios, such as road strength planning in games, this feature of the A-star algorithm is very useful. Because it is not always necessary to obtain the optimal path, for example, the object in the game moves in a safe area and sometimes an approximate optimal solution can be used instead. At this time, you can appropriately sacrifice the accuracy of the algorithm's running results in exchange for the algorithm's running speed.

3. Research on Global Weather Route Planning Method Based on Improved A-Star Algorithm

Global route planning is static planning based on complete prior information. This simulation is using the improved A-star algorithm proposed in this paper to plan global routes for ships. The raster model is used to model the environment of the navigational sea area, and the concave obstacles are convexed and expanded. The simulation proves the efficiency of the improved A-star algorithm proposed in this chapter.

3.1. Selection of Numerical Weather Forecast Information. Wave height is one of the main meteorological factors affecting the navigation of ships. Therefore, the global route planning in this chapter only considers the wave height information in the numerical weather forecast information. In addition, the research in this chapter focuses on the improvement of the A-star algorithm, and for the sake of simplicity, this chapter identifies sea areas with wave heights higher than 6 m as nonnavigable areas like static obstacles (shallow waters, islands, reefs, etc.), and directly the sea area with wave height less than or equal to 6 m [16] is marked as a navigable area that is unobstructed to the navigation of ships.

3.2. Environment Modeling Based on Gird. Because the grid model has a strong ability to express the environment, especially for obstacles with irregular contours and the grid model is simple and easy to operate, this chapter uses the grid model to model the environment. As shown in Figure 5, the starting position of the ship is S, which is the green dot in Figure 5; the target position is E, which is the red dot in the Figure 5; the black shaded area is the static obstacle; and the gray shade is the coverage area is the nonnavigable area formed by meteorological information and is also regarded as a static obstacle.

3.3. Convex Type of Concave Obstacle. As shown in Figure 5, the static obstacles in the upper left and lower left corners are concave obstacles. If the target position is not in the concave area of the concave obstacle, the algorithm's search in the concave area is meaningless. Moreover, the concave area of the concave obstacle may cause some route planning algorithms to fall into local oscillation. Therefore, the concave obstacle needs to be convexed [17]. At the same time, the convex shape of the concave obstacle is also necessary for the improved A-star algorithm proposed in this chapter. As shown in Figure 6, two straight lines are, respectively, tangent to the notches of the obstacle in the upper left corner and the lower left corner to form the line segments ab and cd. The concave obstacle in the upper left corner becomes a convex obstacle surrounded by the line segment ab and the

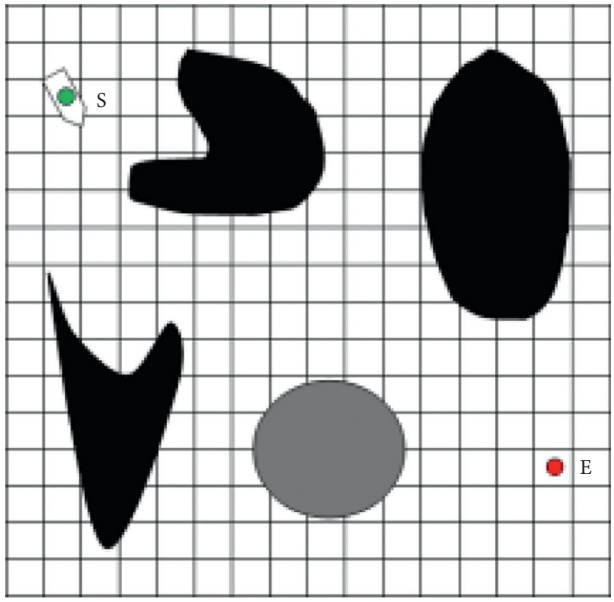


Figure 5: Grid model.

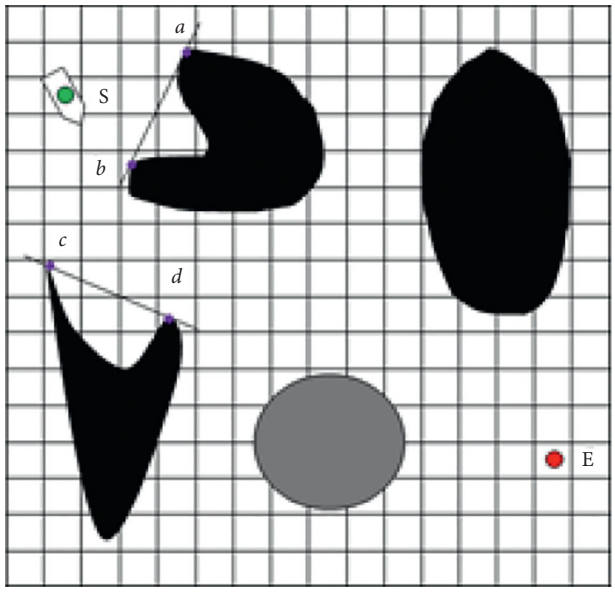


Figure 6: Convex type of concave obstacle.

original obstacle in the upper left corner. The concave obstacle in the lower left corner becomes a convex obstacle surrounded by the line segment *cd* and the original obstacle in the lower left corner.

3.4. The Expansion of Obstacles. As shown in Figure 7, to simplify the problem, the ship is represented by a point, and at the same time, the outline of the obstacle needs to be extended a certain distance. The expansion process in this chapter is based on the convexization of concave obstacles. It is considered that if the grid contains convex obstacles (including convex obstacles formed by the convexity of concave obstacles and those that are not convexized), convex

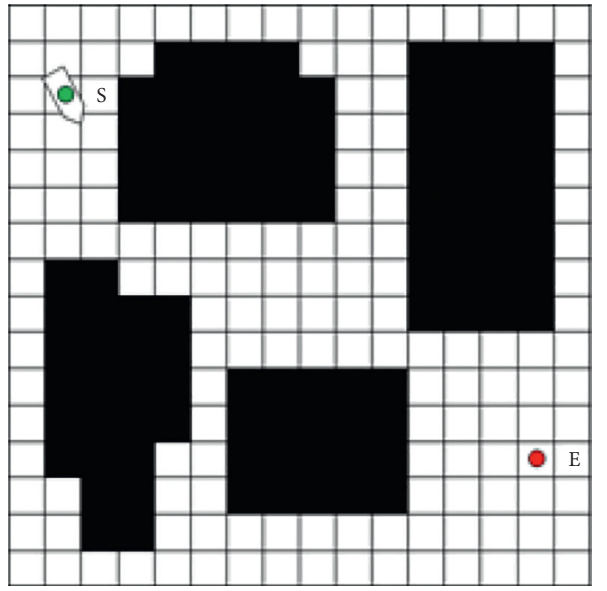


Figure 7: The expansion of obstacles.

barrier is a part of any size, the grid is a nonnavigable area. Because the contour of the convex obstacle generally does not fill the grid, the expansion treatment will also form a certain length of safety distance.

3.5. A-Star Algorithm Improvement. The A-star algorithm is obtained by introducing a heuristic function based on the Dijkstra algorithm, so the classic A-star algorithm is an improved Dijkstra algorithm. Although the classic A-star algorithm has been improved, it still has its shortcomings [18]. For example, in the process of raster map route planning, when the class A-star algorithm is used, it is necessary to traverse all feasible nodes adjacent to each node Once again, when the number of nodes in the raster map is large, the calculation amount of the classic A-star algorithm to find the optimal path will increase sharply, resulting in a decrease in the efficiency of finding the optimal path [19]. The improvement of the A-star algorithm in this chapter is as follows: according to the 0020 position of the target point relative to the starting point, the algorithm's search direction at each node is restricted. Taking a square grid map with 8 search directions as an example, after the preprocessing of the global route planning, the specific improvement process of the A-star algorithm is as follows:

(1) Connect the starting point and the target point with a straight line and measure the angle α between the line and the true north direction. The calculation method of α is as follows:

$$\alpha = \omega \arctan\left(\frac{|\lambda_e - \lambda_s|}{|\phi_e - \phi_s|}\right) + \theta. \tag{5}$$

When $(\lambda_e - \lambda_s) > 0$ and $(\varphi_e - \varphi_s) > 0$, then $\theta = 0^\circ$ and $\omega = 1$.

Table 1: Summary table of the relationship between the α and the direction of the search out-degree.

α	Reserve 5 search out-degree directions	Discard 3 search out-degree directions
[337.5°, 360°) ∪ [337, 5°, 360°)	000T, 045T, 090T, 270T, 315T	135T, 180T, 225T
[22.5°, 67.5°)	000T, 045T, 090T, 135T, 315T	180T, 225T, 270T
[67.5°, 112.5°)	000T, 045T, 090T, 135T, 180T	225T, 270T, 315T
[112.5°, 157.5°)	045T, 090T, 135T, 180T, 225T	270T, 315T, 000T
[157.5°, 202.5°)	090T, 135T, 180T, 225T, 270T	000T, 045T, 315T
[202.5°, 247.5°)	135T, 180T, 225T, 270T, 315T	000T, 045T, 090T
[247.5°, 292.5°)	180T, 225T, 270T, 315T, 000T	045T, 090T, 135T
[292.5°, 337.5°)	225T, 270T, 315T, 000T, 045T	090T, 135T, 180T

Table 2: Comparison of the number of node searches between the classic A-star algorithm and the improved A-star algorithm.

Algorithm	Number of node searches	Percentage reduction in the number of node searches (%)
Classic A-star algorithm	106	—
Improved A-star algorithm	75	29.25

When $(\lambda_e - \lambda_s) > 0$ and $(\varphi_e - \varphi_s) < 0$, then $\theta = 180°$ and $\omega = -1$.

When $(\lambda_e - \lambda_s) < 0$ and $(\varphi_e - \varphi_s) < 0$, then $\theta = 180°$ and $\omega = 1$.

When $(\lambda_e - \lambda_s) < 0$ and $(\varphi_e - \varphi_s) > 0$, then $\theta = 360°$ and $\omega = -1$.

Among them, λ_e is the longitude of the target point, λ_s is the longitude of the starting point, φ_e is the latitude of the target point, φ_s is the latitude of the starting point, and θ and ω are constants.

(2) As shown in Table 1, according to the relationship between the α and the search out degree direction, determine the reserved search out degree direction.

(3) The included angle α is introduced as heuristic information into the A-star algorithm to find the optimal path:

(1) Add the starting point to the open table:

(1) Find the node with the lowest F value in the open table and set the current node to the node with the lowest F value.

(2) Move the current node out of the open table and put it into the closed table.

(3) Determine the 5 nodes pointed to by the reserved 5 search directions in order. (c1) If the node is already in the closed table, then ignore the node. (c2) If the node is not in the open table, then add the node to the open table and use the current node as the parent node of the node and record the F, G, and H values of the node. (c3) If the node is already in the open table, when the node is reached through the current node, and if the value of G is less than the value of G that reached the node without passing through the current node, the current node is taken as the parent node of the node and the F, G, and H values of the node are recalculated. Otherwise, the F, G, and H of the node the value remain unchanged.

(4) Stop searching: if the target point has been added to the closed table, save the path; if the target point is not added to the closed table and the open table is empty, the path does not exist.

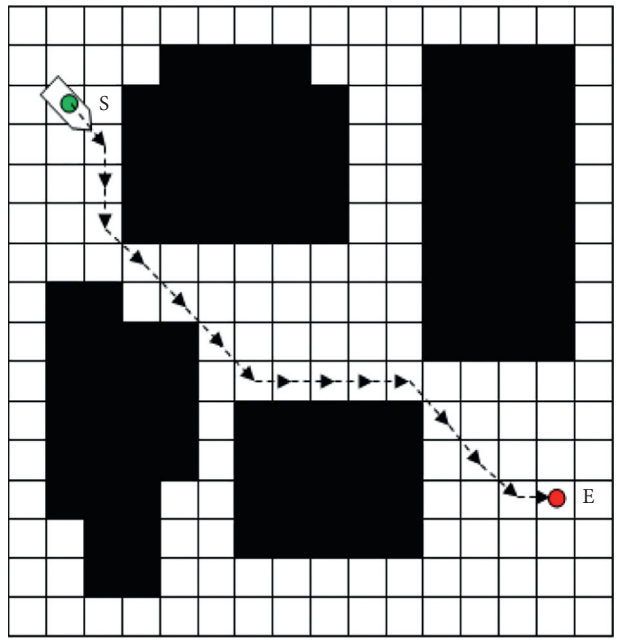


Figure 8: Classic A-star algorithm simulation result graph.

4. Simulation of Global Weather Route Planning Based on Improved A-Star Algorithm

4.1. Simulation content. For the raster map shown in Figure 7, use the classic A-star algorithm and the improved A-star algorithm in this chapter from the starting point S (121.5°E, 38.5°N) to the target point E (134.5°E, 28.5°N). Plan the optimal route and count the number of nodes searched during the route planning process.

For the improved A-star algorithm, according to Equation 5, we can get

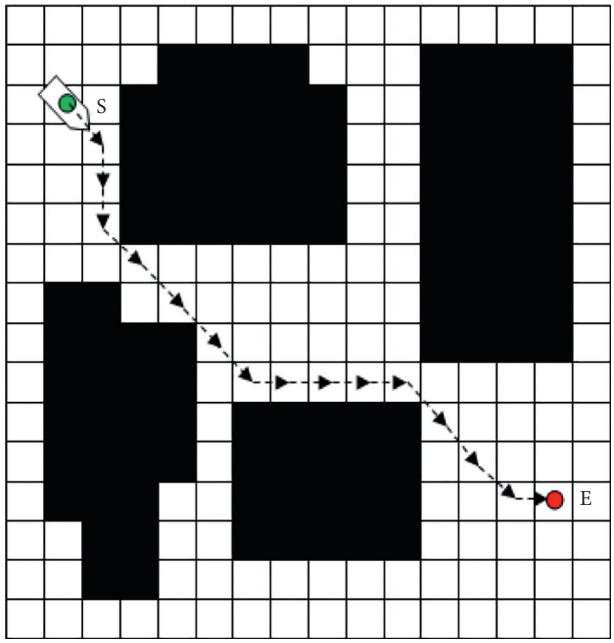


Figure 9: Improved A-star algorithm: algorithm simulation result graph.

$$\alpha = (-1) \times \arctan\left(\frac{134.5 - 121.5}{|28.5 - 38.5|}\right) + 180^{\circ} = 127.57^{\circ} \in [112.5^{\circ}, 157.5^{\circ}). \quad (6)$$

Combined with Table 1, according to the included angle α, the five search out-degree directions that need to be retained are 045T, 090T, 135T, 180T, and 225T.

Therefore, the search direction of each node of the improved A-star algorithm is these 5 directions.

4.2. Simulation Results and Analysis. As shown in Table 2 and Figures 8 and 9, both the classic A-star algorithm and the improved A-star algorithm can find the optimal path. In Figures 8 and 9, compared with the classic A-star algorithm, the improved A-star algorithm reduces the number of node searches by 29.25%, so the improved A-star algorithm proposed in this chapter is feasible and efficient.

5. Conclusion

This paper proposes an improved A-star algorithm for ship weather route planning, in areas where the wave height does not exceed 6 meters and considering static obstacles, the algorithm can find the shortest path, compared with the traditional A∗ algorithm, it reduces the number of search nodes by 29.5%, the time-consuming calculation of the route is reduced, and the efficiency of the algorithm is improved to a certain extent, which can well meet the real-time requirements in the meteorological route planning.

Data Availability

The data used to support the study are not publicly available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funde by the National Key Research (number 2018YFC15007005) and Key R&D projects in Sichuan Province (number 2020YFG0189).

References

[1] Lizhihua and Wanghui, *Marine Weather Navigation*, Dalian Maritime University Press, Dalian, China, (in Chinese), 2006.
[2] Yanglianghua, "Introduction to ship weather navigation," *Meteorological*, vol. 2, no. 7, pp. 19-20, 1982, (in Chinese).
[3] Tangqinghui, *Research on Route Planning Method Based on Electronic Chart*, Ocean University of China, Qingdao, China, (in Chinese), 2011.
[4] H. P. Moravec and A. Elfes, "High resolution maps from angle sonar," in *Proceedings of the International Conference on Robotics Research*, pp. 116–121, St. Louis, MO, USA, March 1985.
[5] R. Chatila and J. Laumond, "Position referencing and consistent world modeling for mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 138–145, Hong Kong, China, March 2014.
[6] J. C. Latombe, "Robot motion planning with uncertainty in control and sensing," *Artificial Intelligence*, vol. 52, no. 52, p. 52, 1991.
[7] A. A. Montes, "Network shortest path application for optimum track ship routing," M. S. Thesis, Operations Research Department Naval Postgraduate School, Monterey, CA, USA, 2005.
[8] C. Huxiaomei and Y. Zhaotao, "Summary of path planning algorithm and its application," *Guizhou: Modern Machinery*, vol. 5, pp. 85–90, 2011, (in Chinese).
[9] R. A. Brooks, *Solving the Find-Path Problem by Good Representation of Free Space. Autonomous Robot Vehicles*, pp. 190–197, Springer, New York, NY, USA, 1990.
[10] J. Wangfengwu, *The Study on the Optimal Ship routing*, no. 2, Dalian Maritime University Press, (in Chinese), pp. 61–64, 1998.
[11] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.
[12] Y. Wang, L. Zhou, B. Yu et al., "3D auto-context-based locality adaptive multi-modality GANs for PET synthesis," *IEEE Transactions on Medical Imaging*, vol. 38, no. 6, pp. 1328–1339, 2019.
[13] D. Harabor and A. Grastien, "The JPS pathfinding system," in *Proceedings of the Fifth Annual Symposium on Combinatorial Search*, Ontario, Canada, July 2012.
[14] P. E. Hart, N. J. Nilsson, and B. Raphael, "Correction to A Formal basis for the heuristic determination of minimum cost paths," *IEEE Transactions on Systems Science & Cybernetics*, vol. 4, no. 2, pp. 100–107, 2007.
[15] S. Koenig and M. Likhachev, "Fast replanning for navigation in unknown terrain," *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 354–363, 2005.
[16] C. Liyuanlin, "Design of optimum ship route using weather routing techniques," *Journal of South China University of Technology (Natural Science)*, vol. 3, 1997.

[17] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 90–98, Taipei, Taiwan, September 2003.

[18] M. Steinbrunn, G. Moerkotte, and A. Kemper, "Heuristic and randomized optimization for the join ordering problem," *The VLDB Journal*, vol. 6, 1997.

[19] P. Tang, C. Zu, M. Hong et al., "DA-DSUnet: dual Attention-based Dense SU-net for automatic head-and-neck tumor segmentation in MRI images," *Neurocomputing*, vol. 435, pp. 103–113, 2021.