Wireless Communications and Mobile Computing

Intelligent Data Management Techniques in Multi-Homing Big Data Networks

Lead Guest Editor: Nawab Muhammad Faseeh Qureshi
Guest Editors: Uttam Ghosh, Varun G. Menon, and Guangjie Han

Intelligent Data Management Techniques in Multi-Homing Big Data Networks

Wireless Communications and Mobile Computing

Intelligent Data Management Techniques in Multi-Homing Big Data Networks

Lead Guest Editor: Nawab Muhammad Faseeh Qureshi
Guest Editors: Uttam Ghosh, Varun G. Menon, and Guangjie Han

Chief Editor

Associate Editors

Academic Editors

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 6168947, 10 pages
https://doi.org/10.1155/2021/6168947

Research Article

Management Information System for Compensation under Multihoming Network Architecture

Cao Yuan [1] **and Hua Yicun** [2]

[1]*Human Resource Department, Donghua University, Shanghai 201620, China*
[2]*The College of Information Science and Technology, Donghua University, Shanghai 201620, China*

Correspondence should be addressed to Hua Yicun; huayicun@dhu.edu.cn

Received 15 April 2021; Revised 27 October 2021; Accepted 5 November 2021; Published 20 November 2021

Academic Editor: Nawab Muhammad Faseeh Qureshi

The management information system for compensation under multihoming network architecture has been developed in order to improve the time efficiency, accuracy, and level of informatization of compensation management in university and deal with the rising data and difficulty in exchanging information between various management information systems resulting from the changing compensation policies. This system is designed based on multihost data network architecture, including function modules of all kinds of compensation promotion, personnel historical data management, time warning, statistics, and report generation. This system integrates my five years' experience in the front-line of compensation management work. The purpose is to fully solve the practical problems of compensation management in universities, truly help the work of compensation administrators, unify the fragmented compensation management works, and comprehensively improve the level of compensation management. It has a strong popularization and signification for reference for the compensation management work of similar universities.

1. Introduction

Compensation information management and statistical analysis is an important routine work of personnel department; the quality of data management directly affects the efficiency and quality of personnel management work [1]. Compensation information has the characteristics of fast update, complex, and trivial. The practical management working process is often faced with serious data redundancy, not timely update, repetitive work, sharing difficulties, and other problems. Information cannot be scientific management, and maintenance [2] will greatly reduce the efficiency and quality of personnel management.

Public institution is a kind of social service organization for the purpose of public welfare in China, which is engaged in education, science and technology, culture, health, and other activities. Most colleges and universities [3] are public institutions, and the compensation management of public institutions has a strong policy focus. In short, the compensation of public institutions is composed of national compensation, uniformly prescribed allowance, local allowance, and institution allowance. The first three parts are uniformly prescribed by national or provincial policies. All public institutions within the scope of policy are subject to uniform references. All national public institutions have the same reference of national compensation, and all public institutions in the same province have the same reference of some provincial allowance. The institution allowance is the only part determined by the institution itself. Therefore, we can find that the compensation of public institutions has a strong continuity, especially the national compensation, which has the same reference all over the country.

For university faculty, all their experiences since they went to university will affect the national compensation and local allowance. Therefore, the compensation management in university is complex. It is necessary to have all the study and work records of each staff to determine their compensation. With the growing number of talents in universities and the rapid development of internet and information technology, compensation management has entered the

digital era, and the informatization construction of universities has been further promoted. As the essential tool of compensation management, a management information system (MIS) for compensation plays a positive role in the management of universities. At the same time, the compensation management in university is characterized by strong policy, large amount of data, and detailed report forms and request highly accurate and safe. Thus, a flexible and efficient MIS for compensation which can make the compensation management work tends to be more scientific, standardized, and modern urgently needs to be developed.

The traditional host network or end users usually only connect to the content source network through a single path. When the only exit path fails, the whole client network cannot get the content source. This problem can be solved by using multihoming technology [4]. The client network can be connected to the outside world through multiple outlets, such as WiFi network [5] and 4G network [6]. When one network fails, it can also transmit data through other networks. In the real network transmission, especially in the big data network [7], the multihoming technology can not only ensure the transmission stability but also improve the transmission efficiency. The client network can simultaneously transmit data through multiple network paths to increase the transmission speed or select the shortest exit of the transmission destination path to reduce the transmission time.

The multihoming technology under TCP/IP architecture [8] has been relatively mature. The classic solutions include host multihoming technology, transport layer multihoming technology, and network layer multihoming technology. Host multihoming technology mainly uses the large address space in IPv6 [9] to communicate with different ISPs depending on multiple IP addresses owned by each host, so as to obtain content from multiple paths. For example, in multiple care of address registration (MCoA [10]), the mobile host or user can register multiple care of addresses for the home address, create multiple binding cache entries, number them with a new binding ID, and send this message in the binding update. Transport layer multihoming technology mainly refers to the design of a new transport protocol supporting multihoming technology to replace TCP technology. For example, Dreibholz et al. designed a new transport layer protocol: flow control transport protocol (SCTP [11]). SCTP is a unicast protocol that supports data exchange between two endpoints and allows each endpoint to have multiple IP addresses. Network layer multihoming technology mainly refers to the technology of inserting a new protocol stack between network layer and transport layer. The representative of network layer multihoming technology is host identity protocol (HIP [12]) and Shim6 [13]. These technologies complete the mapping from a single host name to multiple network addresses by introducing a new namespace, and transition between network layer and transport layer, so as to achieve multihoming transmission.

Based on the practical work of compensation management in our university, this paper takes the digital campus construction [14] as an opportunity to develop a management information system for compensation under multihoming network architecture (MNC-MIS). Based on the in-depth study of the characteristics of compensation management in university and the composition of personnel information, through the analysis and research on the types, structures, and characteristics of our university faculty information, MIS for compensation management is designed, which adapt to the characteristics of university and provided an information platform for university compensation management, analysis, and decision-making. The main contributions are as follows:

(1) The multihoming network architecture of MIS for compensation is designed: this MIS is based on the construction of digital campus. This paper presents the multihoming network architecture of MIS for compensation. Users can obtain compensation management information through LAN, IPv4, IPv6, and 4G

(2) This paper designs a MIS for compensation to adapt to the characteristics of university compensation management: the overall structure of the system is divided into four layers: interaction layer, application layer, data layer, and external layer. In the external layer, the MIS systems that need compensation data support can interact with each other through the unified data bus format. The interaction layer mainly includes the program modules that need to interact with users, which is the entrance and exit of the system. The application layer processes all data management, and the data layer stores system data

(3) The function module design of MNC-MIS is given: including all kinds of function modules of compensation promotion, all kinds of function modules of time warning, function module of personnel history data management, and all kinds of function modules of statistics and report generation. The functions and details of each module are explained

(4) The data flow of MNC-MIS is designed: taking the practical work of new teachers in colleges and universities, the determination and adjustment of social security payment base, the process of monthly compensation issuing and summarizing, and the whole life cycle management as examples, this paper explains the flow and operation of the system data

In the remainder of this paper, problems of the existing compensation data management system are discussed in Section 2. The design of the management information system for compensation under multihoming network architecture is given in Section 3. Section 4 presents the function and effect of MNC-MIS. Finally, the conclusions and future research are provided in Section 5.

2. Problems of the Existing Compensation Data Management System

At present, there are not many colleges and universities that have developed MIS for compensation. There are six major problems of existing compensation management systems.

(1) Data storage and management methods are very backward

The compensation data management of many colleges and universities is mainly composed of paper files and electronic files. These electronic files are also basically electronic documents formed by office software such as word or excel. They are operated on a single machine and can achieve information sharing through U-disk or e-mail, which causes problems such as inaccurate data, repeated processing of information, and difficult to achieve real-time sharing. The existing compensation data management systems in colleges and universities are different, and there is no effective management mode. The commercial development of MIS for compensation is mainly for enterprises and [15], which is expensive, and has full functions but lacks personalized demand, which is difficult to meet the needs of compensation data management in universities.

(2) Salary data cannot be exchanged

The existing compensation management system is self-closed. In the existing system, the common problem is that all kinds of systems are independent of each other and lack of information automatic conversion and sharing function. Operators need to switch back and forth between various systems frequently, which is the inevitable result of the single machine era. Although a large amount of data has been accumulated and collected formally, the repeated storage of these original data has not been scientifically sorted and classified, which brings not only information but also information garbage. Therefore, it is difficult to extract valuable composite information from it.

(3) The workload of information maintenance is large

The information age is full of competition. System developers usually adopt short, flat, and fast system design schemes, lacking overall consideration and optimal combination, resulting in repeated investment of many resources and waste of financial resources, increase the amount of redundant information, occupy storage space, and increase the amount of maintenance. On the other hand, the amount of maintenance comes from the C/S structure of the system itself [16], which is characterized by maintaining both the server and the client.

(4) Lack of big data analysis and auxiliary salary decision-making functions

Management and decision-making are inseparable, but people often ignore the important link of decision-making when developing the system. Compensation management software only provides the daily business, ignoring such important information as the comparison between the growth rate of annual total income and the average compensation level of the society in the same period, the analysis of the salary distribution ratio of all kinds of personnel, the relationship between labour costs and benefits, etc. These data are usually the key to assist decision-making, but they cannot be obtained directly. This is a serious system functional defect.

(5) Lack of automatic reminder function

Current systems usually do not prompt managers what work they should pay attention to, such as a teacher who is close to retirement age and needs to calculate the retirement compensation. Or when the time for unified compensation adjustment is up, you need to modify compensation data in batch. The resulting mistakes often affect the working mood of both sides.

(6) The visualization of management information is poor

At present, the informatization reform of university is still in the exploratory stage, and the degree of management information visualization is low. In the process of information management in universities, problems such as lack of data, noise, and unstructured are often faced, which hinder the information decision-making.

Nowadays, there are many kinds of MIS for compensation management in the market, but most of them are not practical enough, especially the MIS for compensation management in universities. There are many defects such as incomplete function, poor report processing function, complicated query, and statistics. At the same time, the current compensation management system used by our school was developed in the 1990s, which requires low versions of computer hardware and software. Its development background is based on the postcompensation system of the 1993 compensation policy. Obviously, the system cannot adapt to the rapid development of computer science and technology and also cannot meet the requirements of a merit pay system. Therefore, a new MIS for compensation has to be developed.

3. Design of Management Information System for Compensation under Multihoming Network Architecture

3.1. Multihoming Network Architecture. There is a large amount of compensation management data in universities, and the real-time requirement of data processing is high. In this paper, combined with the actual compensation management work of our school, taking the digital campus construction as an opportunity, we develop the MNC-MIS. The schematic diagram of the multihoming data network structure is shown in Figure 1.

The network architecture is divided into three layers: data source, multihoming network, and terminal processor. Among them, the data source stores compensation management data, the multihoming network includes 4G or 5G network [17], LAN [18], IPv4, and IPv6 [19], and the terminal processor can be PC, smartphone, or tablet computer [20]. The advantage of multihoming network architecture is that when one of the network links is disconnected, the end users can still get the data source through other host networks, which will not affect the development of compensation management.

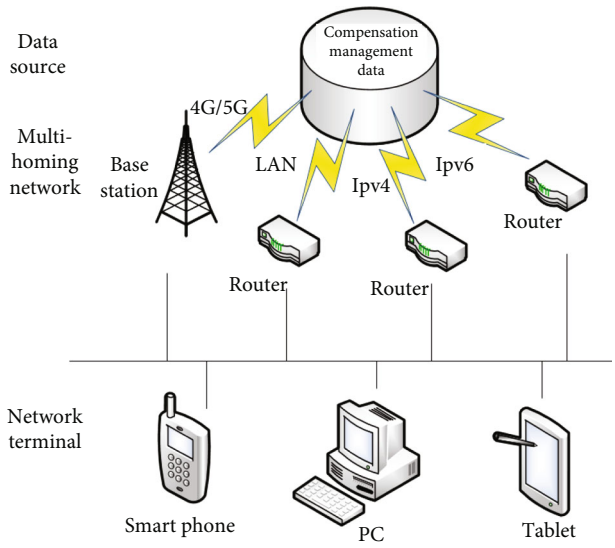


Figure 1: Multihoming network architecture of MNC-MIS.

3.2. Overall Structure Design of MNC-MIS. The overall structure of the system is divided into four layers, the interaction layer, application layer, data layer, and the external layer of other external systems, as shown in Figure 2.

3.2.1. External Layer. Due to the existence of the unified data platform of Donghua University, each MIS can exchange data through the unified data bus format. MIS for personal affairs, finance, and the retired are the systems with the most frequent data interaction with MNC-MIS. These MIS and other MIS of the university which may require compensation data support constitute the external layer.

MIS for personal affairs is mentioned in this layer. Can MNC-MIS directly use personnel information in MIS for personal affairs without having its own personal records database? The answer is no, because although the information in the two MIS is generally called personnel information, the content of the two is actually different. MIS for personal affairs focuses on all kinds of information related to personnel employment and management, while MNC-MIS focuses on all kinds of information related to compensation. Of course, such public information as name, ID, professional title, and position must be reused by both MIS.

3.2.2. Interaction Layer. The interaction layer mainly includes the program modules that need to interact with users, and it is the entrance and exit of the system.

3.2.3. Application Layer. The application layer is the main part of the MIS and the core business logic of the whole system. All the internal processes of data management and processing occur in this layer. At the same time, this layer is the only one that has relations with all the other three layers.

3.2.4. Data Storage Layer. As the name suggests, the data storage layer is where the data is. The core data of this system is divided into two parts: one is the file personal records data and the compensation standard information associated with it. Which data is divided into two parts again, the data of working staff and the retired are stored in the two core databases of the data storage layer. The other core data is the actual payroll data, which also has a special core database. In addition, there are many kinds of subdatabases in this layer, such as historical information database, compensation reference database, and format of report form database. These sublibraries play the role of expanding the core library or supporting the storage of some business logic in the application layer.

3.3. System Function Module Design. On the basis of the overall structure design of the system, we give the system module design, as shown in Figure 3.

The function and significance of some important modules are as follows:

(1) Various compensation promotion modules

As mentioned in the first chapter, the compensation of public institutions has a strong policy focus, according to these policies, such as pay grade salary, Shanghai post allowance and other compensation items will be promoted naturally with the increase of teachers' working years. Of course, promotion cannot simply increase a fixed amount every year but has a series of constraint conditions. For example, promotion of pay grade salary needs to meet the constraints of passing the assessment, being on the job all year round, and not skipping a grade in the previous year, while the promotion of post allowance in Shanghai needs to meet the conditions of seniority being divided exactly by 5, being on the job all year round, and seniority ≤ 35.

All kinds of compensation promotion program modules in this MIS are written to solve the above problems. These modules have embedded judgment programs, which computerize the policy language, so that the computer can calculate the compensation after promotion according to the personal record database and compensation reference database. The working process of these program modules can also be seen in the data flow diagram in the following.

(2) Various time warning modules

In the work of compensation management in universities, due to the large number of teachers in universities, and the existence of dozens of policy which affects the change of compensation of faculty. For example, a new master's or doctor's degree student who has just joined the work will be transferred from an unassigned faculty member to a faculty of certain position level through position orientation after three months. Another example is that a teacher who has reached the age of 60 needs to retire, and his or her on-the-job pay will stop accordingly. There are still many such changes. Therefore, relying on human brain memory or form reminders is not timely or even omitted. The advantages of various time early warning modules in this MIS are reflected here. Through these modules, the system can accurately and timely give the administrator feedback reminder, so that the administrator can have enough time to deal with

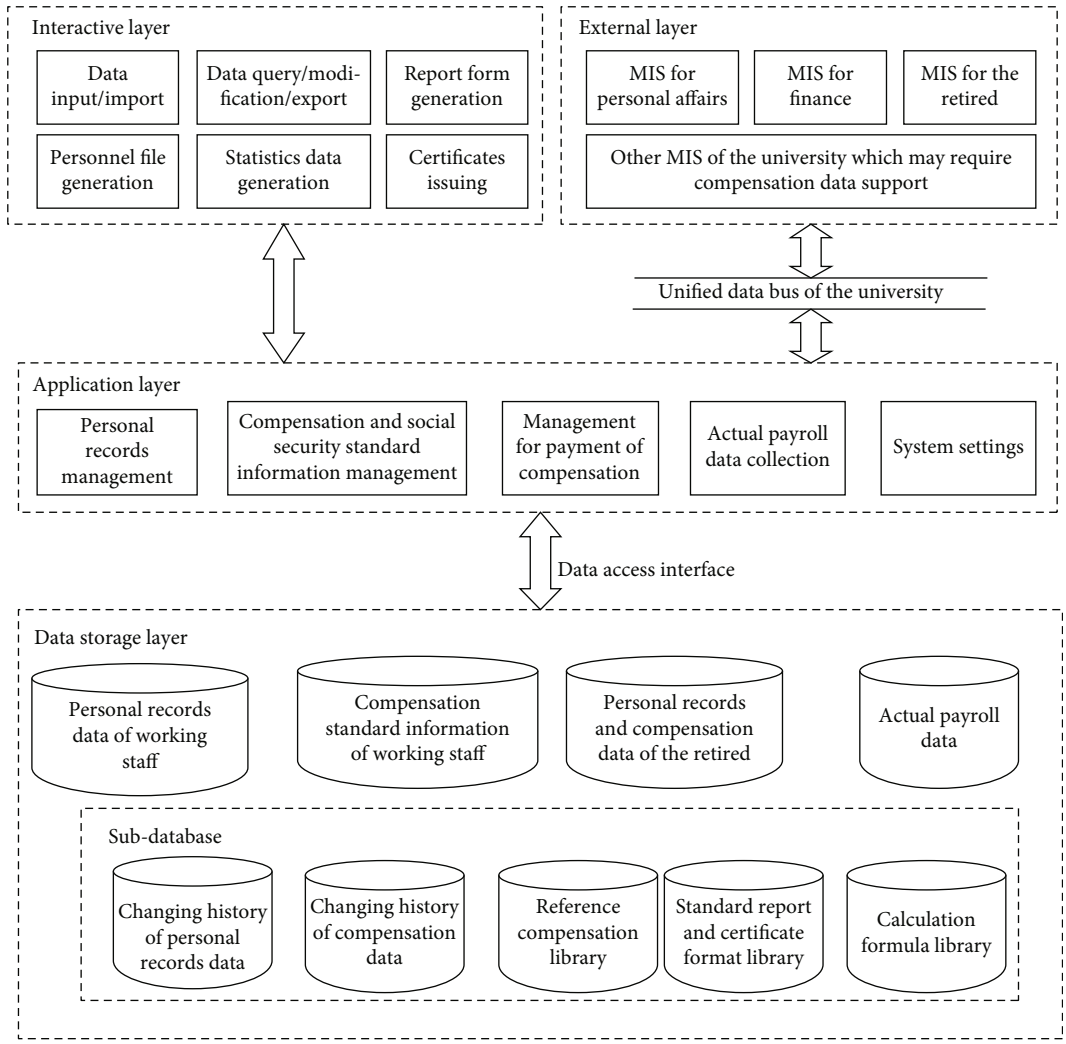


Figure 2: Overall structure of MNC-MIS.

the relevant changes, improve work efficiency, and also enhance the timeliness and accuracy of the work.

(3) Personnel history data management

The compensation policy of public institutions is not only numerous but also requires very meticulous. The whole experience of learning, work, and promotion of faculty is the basis and requirement of compensation promotion of public institutions. For example, in 2006, the starting grade of pay grade salary should be calculated according to the information of each teacher's number of years of schooling, number of years of working, and number of years of being in current position. Therefore, a good MIS for compensation must have the function of historical data record management.

The personnel history data management module shall be able to automatically record each employee's promotion time of title or position, and the compensation adjustment situation every time, and store them in a special database for long-term storage to provide various compensation adjustment policies. With the centralized management of the system, the compensation management has gone away from the original fragmented records, as well as a large number of complicated manual data checking process, and entered an efficient and automatic era.

(4) All kinds of statistics module and report generation module

Compensation in universities should be under the jurisdiction of higher authorities. Take Donghua University as an example, it is under the jurisdiction of Shanghai Municipal Government where it is located, and under the jurisdiction of the Ministry of Education of PRC. The above departments require our university to submit various compensation reports every quarter and every year. However, due to different management departments, these reports not only have different formats but also have different statistical calibers. In the absence of system support, we can only use a lot of time to split and summarize the original data of Excel in different forms according to different requirements, so as to meet the requirements of various reports.

Module for statistics of all kinds of indicators with module for report form generation of this MIS gives a perfect

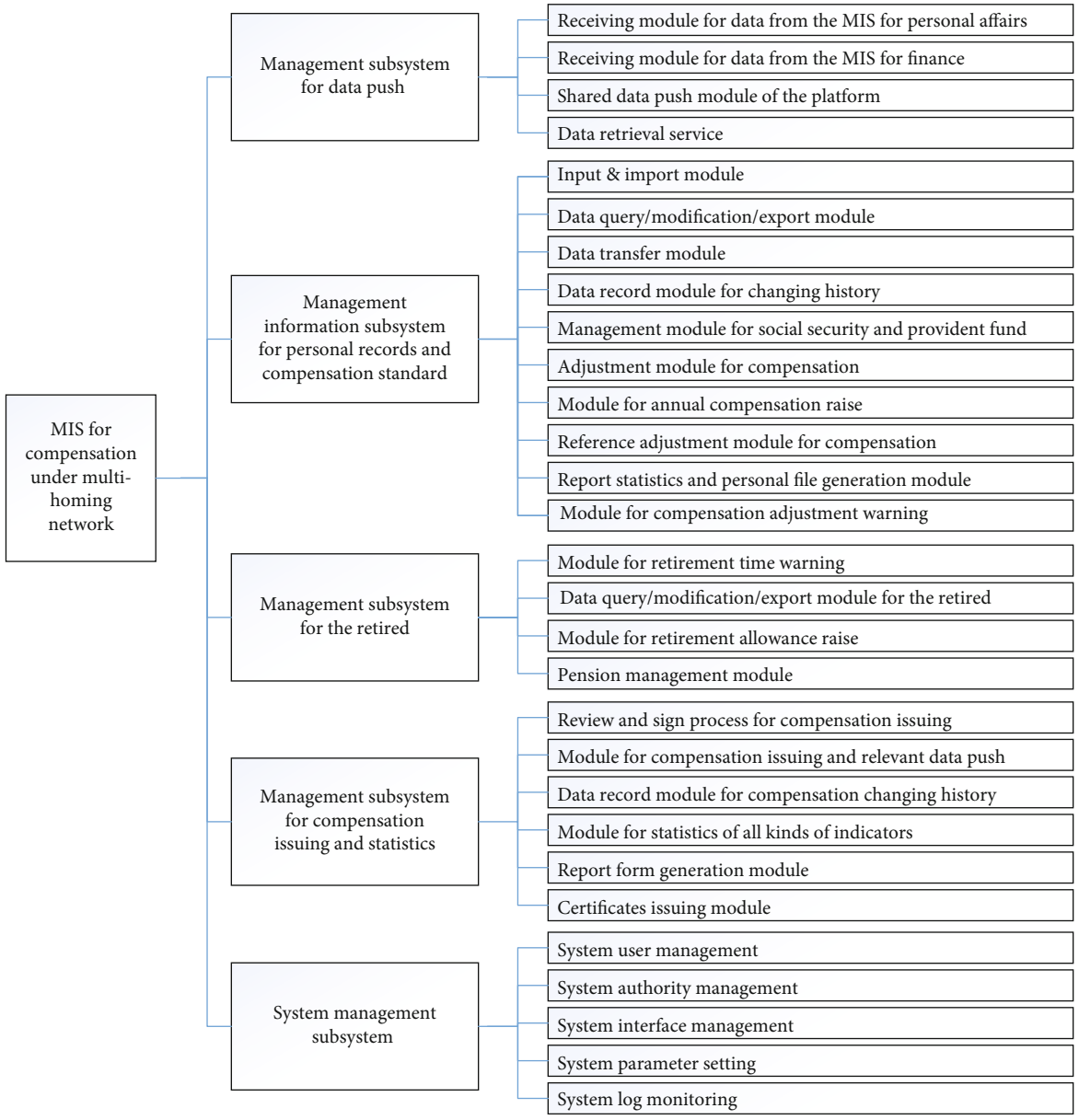


Figure 3: System function module design of MNC-MIS.

solution to this problem. These modules can use the data in both reference compensation database and actual payroll database and give different forms of reports according to the presupposed forms of report. At the same time, the system also has a format editor, which makes it convenient, accurate, and fast to report as you wish.

3.4. Design of System Data Flow. Data flow is the "blood" of each module of the system, so the design quality of data flow also reflects the quality of MNC-MIS and further reflects the level of compensation management in this university. The data flow diagram of MNC-MIS is shown in Figure 4.

The first thing to explain is that the program modules corresponding to entities 18 and 19 in the figure are connected with all the databases in the figure, but if they are lined with all the databases, the overall image will be very chaotic. Therefore, for the convenience of drawing, the "all database in this MIS" database flag represented by dotted line is drawn to refer to all databases in the diagram.

In this figure, we can see not only the operation condition of the data in this MIS but also the process of compensation management in Chinese universities. So, let us take the practical work of several processes in universities to help you understand and read the operation of the system data.

3.4.1. New Teachers Are Coming In. When a new teacher starts his or her career, he or she will first go through the check-in procedures in the personnel section of the human resource department and fill in his or her personal certificate information, study and work experience, and other information in the MIS for personal affairs. The MIS for personal affairs will push the above information related to

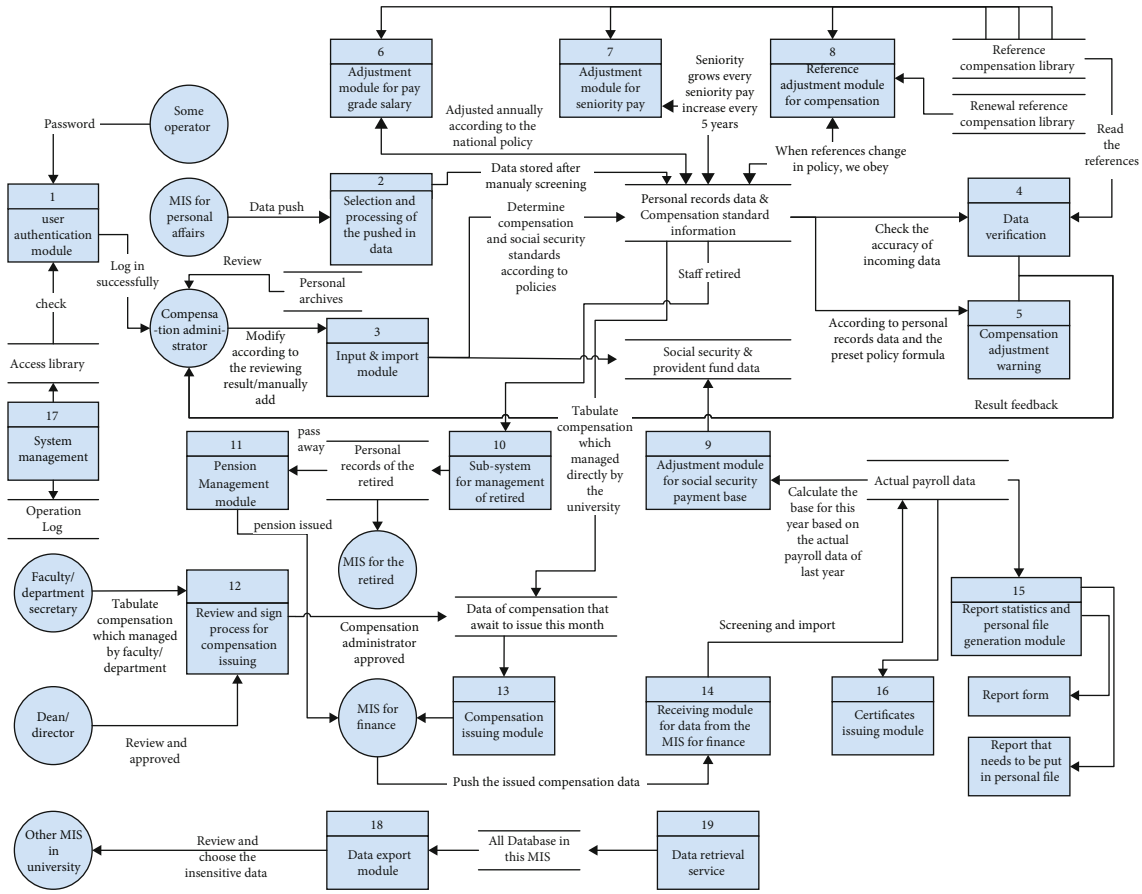


Figure 4: Data flow diagram of MNC-MIS.

determining the compensation standard, and through the selection and processing module shown in entity 2, the selected information will be imported into the personal record database of the MIS. So, how to ensure the accuracy and authenticity of the information? This has to mention the personnel archive system with Chinese characteristics.

Personnel archives are paper materials recording personal identity, education background, qualifications, and other important contents. In China, each unit has a special file management organization. One cannot read its own personnel archive for life, and there are a series of systems to ensure the safety and confidentiality of archives in the process of storage and circulation. Therefore, if we want to determine the accuracy of personal information, we only need to view and refer to the personal file. The compensation administrator has the right to view personal archives.

In Figure 4, you can also see the process of compensation administrator consulting personal archives. According to the results of file reading, manually modify the received push data through entity 3 "input and import module." Then, the administrator should apply the compensation policy according to the above personal file information, manually calculate the standard value of each compensation, and fill in the system. After that, we can call the verification system of entity 4 to verify the compensation standard value. So why not use the computer to calculate the compensation standard in the first place? This is because some items in the compensation cannot be completely calculated by computer and must be judged by comprehensive manual work.

3.4.2. Determination and Adjustment of Social Security Payment Base. Through the above steps, a new employee's monthly compensation standard has been determined. According to this compensation standard, the social security payment base can be calculated. Enter the social security payment base into the social security management program of entity 9, you can get the payment amount of social security and provident fund and store it. At the same time, the social security management program can automatically calculate the social security payment base of all employees in the current year according to the actual payroll data of the previous year and automatically complete the adjustment of the payment amount of the whole staff and the whole project once a year.

3.4.3. Monthly Compensation Issuing and Summarizing Process. There are two main functions of this MIS: one is the management of compensation standard, and the other is the management of monthly compensation issuing process. The system entity 12 module of review and sign process for compensation issuing is the entry of compensation payment process. According to the policy of Donghua University, the national compensation is managed by the compensation administrator of the university, while the school allowance is managed by the faculty/department

and reviewed by the compensation administrator. This approval process is mainly for the school allowance. The faculty/department secretary makes a record of this part of the compensation and initiates the process. After the approval of the dean/director and the approval of the compensation administrator, this part of the compensation will enter into the payroll to be paid this month. In the figure, we can also see that the part of national compensation is directly recorded by the compensation administrator according to the information in the personal compensation standard database, combined with the actual situation of this month, and entered into the compensation database to be paid this month.

After obtaining all the compensation information to be paid this month, through the entity 13 compensation payment and docking module, the compensation information to be paid is pushed through the school unified platform, such as MIS for finance. After that, the compensation is paid to each teacher through the financial system.

Therefore, after the payment is completed, the paid out amount can be imported into the MIS through the closed-loop system, and then, the actual payroll data can be compared with the data of compensation that await to issue. In addition, the actual payroll data is also an important basis for the program modules of entities 9, 15, and 16.

3.4.4. Faculty Life Cycle Management. From the data flow process in Figure 4, we can see that from the beginning of employment to retirement or even death, the school will pay the teachers the relevant treatment in each period. Therefore, the whole life cycle management of employees is also a feature of personnel management in Chinese universities.

When the employee's compensation standard information system is transferred from the on-the-job management system, 10 employees will retire. Because our school has a special retiree management system, so this subsystem mainly plays the role of the link between this MIS and the retirement MIS and can complete some basic retiree information storage and management work. After that, the relevant information of retirees will be pushed to the special retirement MIS through the school platform for the management of the retirement management office.

When a retired teacher dies, the system will read the retiree's information through the entity 11 pension management module, calculate the teacher's pension amount according to the policy, and then transfer it to the financial MIS to complete the pension payment.

4. The Function and Effect of MNC-MIS

The system design fully shows the advantages of computer application in personnel management and provides the system users with as much convenience as possible, such as fast retrieval, comprehensive data, closely following policy, closed-loop management, and personalized operation.

(1) The data format is unified, and the data exchange is accurate and standard

The data format is unified in two aspects:

First, on the aspect of whole school. The system follows the data specification of the unified information platform of the school and can achieve the data docking with other MIS of the school through the unified data bus of the university information platform. The system level data exchange is fast and accurate.

Second, on the aspect of the system itself. There are several main databases and several subdatabases in the system. All databases follow the data protocol agreed in the system. Through the design of this specification, not only the unity of the system data is ensured but also the scalability of the system data is ensured.

(2) The design of data structure emphasizes integrity and comprehensiveness, which meets the needs of compensation management in universities

Data structure design in line with the principle of "can be complete, can be detailed," let the data field be as complete as possible. Data management is based on the principle of more and more, recording every change and every adjustment truthfully and carefully. Because every field, every adjustment, may have an impact on a teacher's subsequent compensation. For example, in some specific cases, two teachers who are consistent in all aspects have only one promotion time difference of one month. This month may bring a gap between the two teachers' salaries in a future compensation adjustment. The emphasis of data integrity and comprehensiveness in this system is the best preparation for such a situation.

(3) To achieve the machine language of compensation policy at all levels and fully achieve the policy adjustment of university compensation

This paper has repeatedly emphasized the characteristics of a strong policy focus of university compensation. Governments at all levels, such as the state, provinces, and cities, have issued a lot of compensation policies, which have been edited into a book and distributed to the compensation administrators of colleges and universities. This system transforms these policies written in natural language into machine language, which means that the computer has learned these policies. Then, the computer can complete all kinds of compensation promotion and adjustment according to these policies.

At the same time, the compensation policy is constantly changing, and the system also retains the interface, allowing the entry of new machine language policy procedures without changing the source code, ensuring the continuity and scalability of the policy.

(4) The closed-loop management of the whole process of compensation management is achieved

The whole process of university compensation management can be closed-loop in this system. From the compensation starting of new employees, to the determination of compensation standard, to the payment of compensation, to the payment of social security and provident fund, to

the summary of actual payroll data, to the review of compensation standard, and finally to statistics and statements. The data of compensation achieves a big close loop. In these large cycles, there are also some small cycles, such as the compensation standard verification process and the social security payment base determination process, which are all small closed-loop cycles in the system. The process of data circulation is the process of compensation management, and the comprehensive informatization of these processes also marks the comprehensive informatization of compensation management, which is the embodiment of efficient, accurate, and professional compensation management in colleges and universities.

(5) Data statistics and report results are instant, convenient, and accurate

Thanks to the unified data format, comprehensive data content, and sufficient policy learning, the process of data statistics and data report generation is efficient and accurate. This process is like a sports car with a well-maintained engine, full oil tank, and shiny body parked in the garage and ready to go. When the driver wants to drive, he or she can go straight away and become a beautiful scenery on the road.

(6) Definable and extensible operation interface, flexible personality

The menu module of MIS supports creation and adjustment, and the operation interface also supports multidimensional configuration. Through the personalized design and adjustment, we can create the most user-friendly interface and give users a good experience.

(7) The factors affecting MNC-MIS

Although MNC-MIS has many advantages, some factors will affect the performance of the system. For example, the bandwidth and security [21] of each network, the performance of the server, the completeness and accuracy of the original salary data, the clarity of salary-related policies, and the speed of policy communication.

To sum up, the MNC-MIS is a system engineering with high design requirements, great development difficulty, and strong technical requirements. This MIS is designed based on my five years' experience in the front-line of compensation management work. The purpose of this MIS is to fully solve the practical problems of compensation management in universities, truly help the work of compensation administrators, unify the fragmented compensation management works, and comprehensively improve the level of compensation management. It has a strong popularization and signification for reference for the compensation management work of similar universities.

5. Conclusions and Future Work

This paper designs a kind of management information system for compensation under multihoming network architecture. Various compensation promotion module, time early warning module, personnel history data management module, statistics module, and report generation module and the detailed data flow have been designed. This MIS can achieve the functions of user authority management, compensation-related information query, information statistics, and system maintenance and solve the problems of not comprehensive function, poor report processing function ability, cumbersome query, and statistics of the existing compensation management system. Using this MIS can achieve efficient intelligent management of compensation data.

In future work, we will strive to improve the big data processing technology of the system, improve the data analysis and processing capacity of the system, improve network performance with advanced methods [22], and adopt or develop advanced artificial intelligence algorithms to further improve the intelligent decision-making and suggestion push function of the system. Moreover, we will seek to exchange information with other universities to establish a general compensation management platform and improve the overall compensation data management ability.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61903078) and 2020 Donghua University Higher Education Planning Special Topic Project (Research on Performance Wage Distribution Method Based on the Sense of Pay Fairness, No. 206-99-0243030).

References

[1] J. Chen, Z. Lv, and H. Song, "Design of personnel big data management system based on blockchain," *Future Generation Computer Systems*, vol. 101, pp. 1122–1129, 2019.

[2] G. Donati and C. Woolston, "Information management: data domination," *Nature*, vol. 548, no. 7669, pp. 613-614, 2017.

[3] J. Hou, Z. Wang, X. Liu et al., "Public health education at China's higher education institutions: a time-series analysis from 1998 to 2012," *Bmc Public Health*, vol. 18, no. 1, p. 679, 2018.

[4] L. Xu, A. Nallanathan, J. Yang, and W. Liao, "Power and bandwidth allocation for cognitive heterogeneous multi-homing networks," *IEEE Transactions on Communications*, vol. 66, no. 1, pp. 394–403, 2018.

[5] K. Poularakis, G. Iosifidis, and L. Tassiulas, "Joint deployment and pricing of next-generation wifi networks," *IEEE Transactions on Communications*, vol. 67, no. 9, pp. 6193–6205, 2019.

[6] V. Kumar and N. B. Mehta, "Modeling and analysis of differential cqi feedback in 4g/5g ofdm cellular systems," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2361–2373, 2019.

[7] A. D'Alconzo, I. Drago, A. Morichetta, M. Mellia, and P. Casas, "A survey on big data for network traffic monitoring and analysis," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 800–813, 2019.

[8] S. Babu, A. Rajeev, and B. S. Manoj, "A medium-term disruption tolerant SDN for wireless TCP/IP networks," *IEEE Transactions on Network and Service Management*, vol. 17, no. 4, pp. 2318–2334, 2020.

[9] L. K. Indumathi and D. S. Punithavathani, "Performance improvement of proxy mobile ipv6 for the support of multihoming," *Wireless Personal Communications: An International Journal*, vol. 96, no. 2, pp. 1653–1672, 2017.

[10] I. Cho, K. Okamura, T. W. Kim, and C. S. Hong, "Performance analysis of IP mobility with multiple care-of addresses in heterogeneous wireless networks," *Wireless Networks*, vol. 19, no. 6, pp. 1375–1386, 2013.

[11] T. Dreibholz, E. P. Rathgeb, I. Rüngeler, R. Seggelmann, M. Tüxen, and R. R. Stewart, "Stream control transmission protocol: past, current, and future standardization activities," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 82–88, 2011.

[12] A. Keranen, "This document specifies two transport modes for host identity protocol (HIP) signaling messages that allow conveying them over encrypted connections initiated with the host identity protocol," *Radiology*, vol. 98, no. 98, pp. 605–610, 2015.

[13] Y. Qiao, E. Fallon, J. Murphy, L. Murphy, Z. Shi, and A. Hanley, "Transmission scheduling for multi-homed transport protocols with network failure tolerance," *Telecommunication Systems*, vol. 43, no. 1-2, pp. 39–48, 2010.

[14] J. Cao, Z. Li, Q. Luo, and Q. Hao, "Research on the construction of smart university campus based on big data and cloud computing," in *2018 International Conference on Engineering Simulation and Intelligent Control (ESAIC)*, pp. 351–353, Hunan, China, 2018.

[15] K. Melendez, A. Dávila, and M. Pessoa, "Information technology service management models applied to medium and small organizations: a systematic literature review," *Computer Standards & Interfaces*, vol. 47, pp. 120–127, 2016.

[16] L. Zhong, "Monitoring function design of radio monitoring management system based on C/S architecture," in *2019 10th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 424–428, Qingdao, China, 2019.

[17] M. Agiwal, H. Kwon, S. Park, and H. Jin, "A survey on 4G-5G dual connectivity: road to 5G implementation," *IEEE Access*, vol. 9, pp. 16193–16210, 2021.

[18] W. Kim, J. Park, J. Jo, and H. Lim, "Covert jamming using fake ACK frame injection on IEEE 802.11 wireless LANs," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1502–1505, 2019.

[19] J. Xie and U. Narayanan, "Performance analysis of mobility support in IPv4/IPv6 mixed wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, pp. 962–973, 2010.

[20] J. Q. Zhu, Y. L. Ban, Y. Zhang et al., "A novel capacitive coupler array with free-positioning feature for mobile tablet applications," *IEEE Transactions on Power Electronics*, vol. 34, no. 7, pp. 6014–6019, 2019.

[21] H. N. Noura, R. Melki, M. Malli, and A. Chehab, "Lightweight and secure cipher scheme for multi-homed systems," *Wireless Networks*, 2020.

[22] A. AI-Najjar, F. H. Khan, and M. Portmann, "Network traffic control for multi-homed end-hosts via SDN," *IET Communications*, vol. 14, no. 19, pp. 3312–3323, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 2888673, 20 pages
https://doi.org/10.1155/2021/2888673

Review Article

The Influence of Big Data Analytics on E-Commerce: Case Study of the U.S. and China

Weiqing Zhuang

School of Internet Economics and Business, Fujian University of Technology, Fuzhou, China

Correspondence should be addressed to Weiqing Zhuang; zmakio@aliyun.com

Received 27 July 2021; Revised 3 October 2021; Accepted 21 October 2021; Published 31 October 2021

Academic Editor: Varun Menon

Big data analytics (BDA) is a wide and deep application in e-commerce, which impacts positively on the global economy, especially the U.S. and China who have done well. This paper seeks to examine the relative influence of theoretical research and practical activities of BDA in e-commerce to explain the differences between the U.S. and China according to the two main literature databases, Web of Science and CNKI, respectively, and by employing other samples that present retail e-commerce sales and the number of some data companies founded in the U.S. and China each year. We further determine the reasons leading to the difference between the U.S. and China in BDA in e-commerce, which can help managers devise appropriate business strategies in e-commerce for each of them, and provide a proof of the significant relationship of theoretical research and practical activities in BDA in e-commerce. In addition, the variables related to big data companies show a moderation effect rather than mediating effect relative to the practice of theoretical research in e-commerce in the United States, but they show a moderate effect and mediating effects in China. The results of this study help clarify doubts regarding the development of China's e-commerce. Moreover, three orientations in e-commerce using BDA and the use of quantum computing in e-commerce to solve existing e-commerce problems are explored to provide better evidence for decision-making that could be valuable in future research.

1. Introduction

Big data are a frontier topic for researchers [1] and have always influenced academic research [2]. However, there are six debates regarding the aspects of the big data literature, including different approaches to big data analytics (BDA), artificial intelligence, big data capability, big data-driven business organizing, big data access, and social risks of big data value realization [3]. The severer is the privacy trust crisis in the era of big data, both in the field of enterprise services and some public services; moreover, there are also doubts about data since individuals may have inaccurate information [4]. More precisely, the key point of the debate and crisis regarding big data is how to use big data well and in a correct manner that is frank and honest, agreeing at the outset to focus only on what really matters [5]. For example, considerations include two things: how to adjust algorithms to ever-changing conditions [6] and how to evade the negative effect of "big data hubris" [7].

Hence, general statistical techniques and computational algorithms are issues that require different tools to manage big data sets [8], particularly with respect to different qualities of information disclosure for different purposes. BDA requires understanding the relationships among features and the explored data [9]. It has evolved from the statistical techniques for data mining from the 1970s to business intelligence (BI) 3.0 today [10]. Another reason that prompts us to focus on this area is that China accounted for 23.1 percent of the total online retail sales in 2017, while the United States ranked behind the UK, South Korea, and Denmark, only sharing 9 percent. Data come from the Statista website, https://www.statista.com/statistics/255083/online-sales-as-share-of-total-retail-sales-in-selected-countries/. Obviously, e-commerce marketing not only depends on information

technology or the population, but it is likely to reflect on the oneness of value and perception that the U.S. lacks. Therefore, the situation of the United States and China regarding research on and practice of BDA in e-commerce is worthy of discussion.

Akter and Wamba [11] fully demonstrate that BDA in e-commerce, as an emerging field since 2006, exhibits strategy-led analytics and has sustainable value-driven facets for businesses involved in organization management, goods sales, production management, data quality, IT infrastructure and its security, HRM, overarching values, and so on. Additionally, Manyika et al. [12] propose five major contributions that big data can promote to businesses: creativity, performance, consumer behavior, decision-making, and innovative business models. In short, a series of new issues in BDA in e-commerce should be focused on, including how to determine the first-rank relationship among a commodity's dynamic pricing, dynamic subsidizing, and cost to e-commerce parties, and how to enact good policies with BDA in dynamic pricing to buyers that depend on a buyer's past good or bad actions, including comments, sales returns, and sharing. Actually, goods prices have not been adjusted to a per-buyer basis in e-commerce firms, while their expenditures differ in completing the transaction. These aspects require a determination of whether there are any respective methods and experiences for the U.S. and China to solve these issues. This paper summarizes the relevant literature and applications as follows.

1.1. U.S. Similarly, there are also many examples that occur in practice. Data breaches occur daily in the American society; Taylor [13] provides a list of 17 breaches, such as the Yahoo 500 million user accounts attacked in 2013, Adult Friend Finder in 2016, eBay in 2014, Equifax in 2017, Heartland Payment Systems in 2008, Target Stores in 2013, TJX Companies in 2006, Uber in 2016, JP Morgan Chase in 2014, US OPM in 2012, Sony's PlayStation Network in 2011, Anthem in 2015, RSA Security in 2011, Stuxnet in 2010, VeriSign in 2010, Home Depot in 2014, and Adobe in 2013. Of course, all of these occurrences are just the tip of the iceberg, as Facebook's data crisis has made Americans worried about and disgusted with big data once again. To some extent, Americans pay more attention to the safety and rationality of data use may be due to their cultural and cognitive characteristics [14], which is somewhat different from China [15, 16].

1.2. China. By first searching "big data" and then add "business" to the Chinese literature database "CNKI" to classify papers by title, only 34 papers were found, and no paper was retrieved when also searching for "crisis" in these 34 papers. Moreover, some critics state that big data in China are overhyped because companies are more interested in using big data to attract media and investors. Data come from the China government website; "How Baidu, Tencent, and Alibaba are leading the way in China's big data revolution," http://www.scmp.com/tech/innovation/article/1852141/how-baidu-tencent-and-alibaba-are-leading-way-chinas-big-data. However, increasing cases of stealing and trafficking of personal information have been reported in China, covering hundreds of millions of items in transportation, logistics, health care, social networking, and banking. For example, 300 million user information leaks from Shunfeng Express and 500 million user data leakages from Huazhu Group or Wanhao Group have occurred.

Regardless of how serious the criticisms are, a notable example shows that Alibaba predicted the dynamic change of income from impoverished people through mining and analyzing their transaction data on its e-commerce platform and then helped the Chinese government to target poverty alleviation. Of course, since this process is not accurate enough to identify each impoverished person's income, Alibaba has cooperated with the government to develop a platform of "Internet plus targeted poverty alleviation," which connects several public service data and also clarifies the main reason why certain people are still poor, including illness that leads to poverty or unemployment, disasters, laziness, etc.

This paper is composed of four sections and focuses on two topics: BDA in e-commerce for the U.S. and China in academic research and in practice. The second section presents a discussion of empirical analysis on the connection of BDA in e-commerce in theory and practice for the United States and China. The third section (Conclusions) sums up the comparison of BDA in e-commerce between the U.S. and China in theoretical research and in practice. The last section explores avenues for future research and application. This paper focuses on comparing BDA in e-commerce between the United States and China by examining their level of theoretical research and practical application by quantifying the literature and the market and on explaining the status of e-commerce development and the main factors that influence it in the U.S. and China.

2. Discussion

There are several issues, such as the relationship between academic researches and practical activities of BDA in e-commerce, the difference in the relationship between the U.S. and China, and how academic research affects practical activity on BDA in e-commerce for U.S. and China severally, all of which should be discussed in depth after detailed literature review and the analysis above. It widely adopts a statistical analysis method for investigating the effect that theoretical research promotes dramatically practical activities in BDA in e-commerce. Specifically, what follows builds regression models of academic research and its application to find out the action mechanism of BDA applied in e-commerce and then has a comparison of the moderating and mediating effects of BDA between U.S. and China and has an investigation on lag consideration of academic research response to practical application in BDA for e-commerce.

2.1. Data Acquisition. The relevant literature for BDA in e-commerce between U.S. and China, collected by two main literature databases, Web of Science and CNKI, respectively, is aimed at U.S. and China, during the period from 1990 to

Table 1: Retail e-commerce sales in the U.S. and China.

Year	U.S.				China					
	Total retail sales ($ billion)	Retail e-commerce sales ($ billion)	Proportion of e-commerce in total retail	Growth rate of retail e-commerce	Total retail sales (¥ billion)	Retail e-commerce sales (¥ billion)	Total retail sales** ($ billion)	Retail e-commerce sales** ($ billion)	Proportion of e-commerce in total retail	Growth rate of retail e-commerce
2017	5076.00	453.46	8.93%	16.54%	36626.20	7175.10	5420.64	1061.91	19.59%	39.17%
2016	4856.33	389.11	8.01%	14.39%	33231.63	5155.57	5002.46	776.08	15.51%	32.97%
2015	4725.99	340.16	7.20%	13.98%	30093.08	3877.32	4789.13	617.05	12.88%	38.98%
2014	4639.44	298.44	6.43%	14.47%	27189.61	2789.80	4415.24	453.03	10.26%	49.70%
2013	4458.45	260.72	5.85%	13.23%	23780.99	1863.60	3865.39	302.91	7.84%	42.15%
2012	4302.23	230.26	5.35%	15.38%	21030.70	1311.00	3332.67	207.75	6.23%	67.52%
2011	4102.95	199.56	4.86%	17.45%	18391.86	782.60	2844.60	121.04	4.26%	53.72%
2010	3818.05	169.92	4.45%	16.78%	15699.84	509.10	2319.04	75.20	3.24%	96.87%
2009	3612.47	145.51	4.03%	2.76%	13267.84	258.60	1942.32	37.86	1.95%	105.73%
2008	3935.32	141.59	3.60%	3.75%	10848.77	125.70	1560.35	18.08	1.16%	124.06%
2007	3995.18	136.47	3.42%	20.41%	8921.00	56.10	1171.77	7.37	0.63%	163.38%
2006	3871.57	113.33	2.93%	24.02%	7641.00	21.30	958.37	2.67	0.28%	10.36%
2005	3689.28	91.39	2.48%	25.88%	6717.66	19.30	819.73	2.36	0.29%	141.25%
2004	3473.05	72.60	2.09%	27.02%	5950.10	8.00	719.16	0.97	0.13%	105.13%
2003	3262.73	57.16	1.75%	28.08%	5251.63	3.90	634.47	0.47	0.07%	129.41%
2002	3128.55	44.62	1.43%	30.25%	4813.59	1.70	581.57	0.21	0.04%	183.33%
2001	3062.27	34.26	1.12%	24.09%	4305.54	0.60	520.06	0.07	0.01%	—
2000	2983.28	27.61	0.93%	90.53%	3910.57	—	472.37	—	—	—
1999	2803.09	14.49	0.52%	190.73%	3564.79	—	430.64	—	—	—
1998	2581.76	4.98	0.19%	—	3337.81	—	402.11	—	—	—

Notes: (1) the data from 2017 U.S. total retail sales and retail e-commerce sales are from the Digitalcommerce360 website (https://www.digitalcommerce360.com/article/us-ecommerce-sales/); other years are from the United States Census Bureau website (https://www.census.gov/); the data for 2017 China total retail sales and retail e-commerce sales are from the China National Bureau of Statistics website and ECRC website (http://www.stats.gov.cn/tjsj/zxfb/201802/t20180228_1585631.html and http://www.100ec.cn/zt/17wlls/); other years are from Yue Hongfei, National Report on E-Commerce Development in China. Inclusive and Sustainable Industrial Development Working Paper Series WP 17, 2017. United Nations Industrial Development Organization. (2) **The yearly average exchange rate for USD (US Dollar) to CNY (Chinese Yuan) during the period from 1990 to 2017 shown on the website https://www.ofx.com/en-us/forex-news/historical-exchange-rates/yearly-average-rates/. (3) Date (GMT), (4) rate, (5) date (GMT), (6) rate, (7) date (GMT), (8) rate, (9) 31-Dec-2017, (10) 6.756806, (11) 31-Dec-2007, (12) 7.613239, (13) 31-Dec-1997, (14) 8.319331, (15) 31-Dec-2016, (16) 6.643058, (17) 31-Dec-2006, (18) 7.972895, (19) 31-Dec-1996, (20) 8.338875, (21) 31-Dec-2015, (22) 6.283627, (23) 31-Dec-2005, (24) 8.19495, (25) 31-Dec-1995, (26) 8.370025, (27) 31-Dec-2014, (28) 6.158134, (29) 31-Dec-2004, (30) 8.273679, (31) 31-Dec-1994, (32) 8.639665, (33) 31-Dec-2013, (34) 6.152292, (35) 31-Dec-2003, (36) 8.277176, (37) 31-Dec-1993, (38) 5.779529, (39) 31-Dec-2012, (40) 6.310468, (41) 31-Dec-2002, (42) 8.276877, (43) 31-Dec-1992, (44) 5.52057, (45) 31-Dec-2011, (46) 6.46553, (47) 31-Dec-2001, (48) 8.278869, (49) 31-Dec-1991, (50) 5.333729, (51) 31-Dec-2010, (52) 6.769961, (53) 31-Dec-2000, (54) 8.278676, (55) 31-Dec-1990, (56) 4.792069, (57) 31-Dec-2009, (58) 6.830938, (59) 31-Dec-1999, (60) 8.277917, (61) 1 unit of USD = X units of CNY, (62) 31-Dec-2008, (63) 6.952764, (64) 31-Dec-1998, 8.300753.

Table 2: Listing of a limited number of data companies in the U.S. and China by founded year.

Founded year	Number of data companies in the U.S.					Number of data companies in China				
	Total	Data/ technology	Business analytics	Industrial application	Research/ consulting	Total	Data/ technology	Business analytics	Industrial application	Research/ consulting
2016						106	13	17	40	36
2015	1	0	0	0	1	221	35	31	128	27
2014	8	2	0	6	0	230	34	35	140	21
2013	30	8	3	18	1	148	27	28	80	13
2012	41	5	3	32	1	129	31	19	71	8
2011	51	9	1	37	4	110	23	25	58	4
2010	50	10	4	35	1	79	20	10	46	3
2009	32	7	1	23	1	76	18	10	43	5
2008	26	5	2	14	5	66	15	17	28	6
2007	28	8	1	18	1	57	16	17	22	2
2006	21	3	1	15	2	49	11	13	20	5
2005	17	2	1	14	0	47	15	7	22	3
2004	13	2	0	10	1	30	9	3	18	0
2003	11	4	2	5	0	42	18	3	20	1
2002	6	1	0	5	0	27	7	6	12	2
2001	10	1	2	3	4	30	11	1	17	1
2000	16	6	1	7	2	38	13	6	17	2
1999	10	3	1	6	0	22	9	5	7	1
1998	10	1	1	7	1	19	7	2	10	0
1997	4	0	0	4	0	12	2	4	6	0
1996	4	1	0	3	0	1	1	0	0	0
1995	6	2	0	4	0	10	3	4	3	0
1994	4	0	1	3	0	3	2	0	1	0
1993	5	1	0	2	2	4	1	1	2	0
1992	2	2	0	0	0	4	2	0	2	0
1991	0	0	0	0	0	0	0	0	0	0
1990	2	0	1	1	0	0	0	0	0	0
1980s	27	3	3	17	4	3	2	0	1	0
1970s	18	4	6	6	2	0	0	0	0	0
1960s	16	4	1	5	6	0	0	0	0	0
1950s	5	1	2	2	0	2	1	0	0	1
Total	474	95	38	302	39	1565	346	264	814	141

Notes: the data of the number of data companies in the U.S. are from the OpenData500 website (http://www.opendata500.com/us/list/), and China's data are from the Data Technology Industry Innovation Institute (a report published a listing of 1574 big data companies in 2017); the category of data companies both in the U.S. and China is definitely displayed according to their initial data.

2017. As a whole, we present the results of three stages of searching for subject terms classified by title from several literature databases.

Table 1 presents the growth of retail e-commerce sales in the U.S. and China, and it is observed that the speed of development in the U.S. is significantly different from that in China. China's e-commerce market saw high growth in he past and in the present which will continue to rise in the future. It will lead to a tremendous market in the e-commerce industry applying big data compared with the U.S.

Table 2 lists a certain (limited) number of data companies in the U.S. and China divided by year, which is subject to the difficulty of obtaining the complete information of big data companies. This table collects and organizes data from "OpenData500" for the U.S. and "Data Technology Industry Innovation Institute" for China and can be used to perform a correlation analysis on the issue of theoretical research in BDA in e-commerce to guide practices for the U.S. and China.

2.2. Variable Setting and Data Disposal. Conducting an empirical analysis of the connection of BDA in e-commerce in theory and practice, such as in comparing the United States and China, requires variables to be set up to represent various subjects for literature retrieval and practical activities. The details are shown in Table 3. For instance, searching the subject term of "E-Commerce" in

Table 3: Variable setting.

(a)

Searching subject term	Literature database	Variable	Searching subject term	Literature database	Variable	Searching subject term	Literature database	Variable	Searching subject term	Literature database	Variable
"E-Commerce"	ProQuest	X01	"Business Intelligence Analytics"	WoS	X11	"Apriori"	WoS	X21	"Clustering Algorithm" and "Big Data"	WoS	X31
	WoS	X02		CNKI	X12		CNKI	X22		CNKI	X32
	CNKI (all)	X03	"Big Data Model"	WoS	X13	"*k*-means" and "Big Data"	WoS	X23	"Cloud" and "Big Data"	WoS	X33
	CNKI	X04		CNKI	X14		CNKI	X24		CNKI	X34
	CNKI (master's & doctoral dissertation)	X05	"Big Data Algorithm"	WoS	X15	"SVM" and "Big Data"	WoS	X25	"Regression" and "Big Data"	WoS	X35
	ProQuest	X06		CNKI	X16		CNKI	X26		CNKI	X36
"E-Commerce" and then "Big data"	WoS	X07	"Hadoop"	WoS	X17	"Machine Learning" and "Big Data"	WoS	X27	"Decision Analytics" and "Big Data"	WoS	X37
	CNKI (all)	X08		CNKI	X18		CNKI	X28		CNKI	X38
	CNKI	X09	"MapReduce"	WoS	X19	"Deep Learning" and "Big Data"	WoS	X29	"Optimization" and "Big Data"	WoS	X39
	CNKI (master's & doctoral dissertation)	X10		CNKI	X20		CNKI	X30		CNKI	X40
"Genetic Algorithm" and "Big Data"	WoS	X41	"Classifier" and "Big Data"	WoS	X49	"Online Consumer Behavior"	WoS	X57	"Cloud Computing & E-Commerce"	WoS	X65
	CNKI	X42		CNKI	X50		CNKI	X58		CNKI	X66
"Neural Networks" and "Big Data"	WoS	X43	"Social Network" and "Big Data"	WoS	X51	"Online Consumer Behavior & Big Data"	WoS	X59	"Artificial Intelligence & Big Data & E-Commerce"	WoS	X67
	CNKI	X44		CNKI	X52		CNKI	X60		CNKI	X68
"Text Analysis" and "Big Data"	WoS	X45	"Prediction Model" and "Big Data"	WoS	X53	"Internet of Things & E-Commerce"	WoS	X61	"Quantum Computing"	WoS	X69
	CNKI	X46		CNKI	X54		CNKI	X62		CNKI	X70
"Association Rules" and "Big Data"	WoS	X47	"E-commerce & Big Data Analytics"	WoS	X55	"Mobile Technology & E-Commerce"	WoS	X63			
	CNKI	X48		CNKI	X56		CNKI	X64			

(b)

Object	Content	Variable	Object	Content	Variable	Object	Classification	Variable	Object	Classification	Variable
U.S.	Total retail sales	Y01	China	Total retail sales	Y05	Founded number of data companies in the U.S.	Total	Y09	Founded number of data companies in China	Total	Y14
	Retail e-commerce sales	Y02		Retail e-commerce sales	Y06		Data/technology	Y10		Data/technology	Y15
	Proportion of e-commerce in total retail	Y03		Proportion of e-commerce in total retail	Y07		Business analytics	Y11		Business analytics	Y16
	Growth rate of retail e-commerce	Y04		Growth rate of retail e-commerce	Y08		Industrial application	Y2		Industrial application	Y17
							Research/consulting	Y13		Research/consulting	Y18

Notes: the default of "WoS" stands for "WoS (Core Collection)" and "CNKI" stands for "CNKI (periodical)."

Table 4: Partial listing of retail e-commerce sales in the U.S. and China.

Year	U.S.				China			
	Total retail sales ($ billion)	Retail e-commerce sales ($ billion)	Proportion of e-commerce in total retail	Growth rate of retail e-commerce	Total retail sales (¥ billion)	Retail e-commerce sales (¥ billion)	Proportion of e-commerce in total retail	Growth rate of retail e-commerce
2001	3062.27	34.26	1.12%	24.09%	4305.54	0.60	0.01%	154.54%
2000	2983.28	27.61	0.93%	90.53%	3910.57	0.235719337	$6.02775E-05$	162.54%
1999	2803.09	14.49	0.52%	190.73%	3564.79	0.089784161	$2.51864E-05$	170.55%
1998	2581.76	4.98	0.19%	75.26%	3337.81	0.033185792	$9.94239E-06$	178.56%
1997	2578.32	2.84149264	0.110%	79.68%	3125.29	0.011913337	$3.81191E-06$	186.56%
1996	2460.17	1.581418432	0.064%	84.10%	2836.02	0.004157362	$1.46591E-06$	194.57%
1995	2342.02	0.858999692	0.037%	88.52%	2361.38	0.001411333	$5.97673E-07$	202.58%
1994	2223.87	0.455654409	0.020%	92.95%	1862.29	0.000466433	$2.50462E-07$	210.58%
1993	2105.72	0.236151546	0.011%	97.37%	1427.04	0.000150181	$1.0524E-07$	218.59%
1992	1987.57	0.11964916	0.006%	101.79%	1099.37	$4.71393E-05$	$4.28785E-08$	226.60%
1991	1869.42	0.059293899	0.003%	106.21%	941.56	$1.44334E-05$	$1.53292E-08$	234.60%
1990	1751.27	0.028754134	0.002%	110.63%	830.01	$4.31362E-06$	$5.19706E-09$	242.61%

Notes: the underlined values are simulated values.

Table 5: Partial listing of a limited number of data companies in the U.S. and China by founded year.

Founded year	Number of data companies in the U.S.					Number of data companies in China				
	Total	Data/ technology	Business analytics	Industrial application	Research/ consulting	Total	Data/ technology	Business analytics	Industrial application	Research/ consulting
2017	34	7	2	24	2	154	29	26	83	17
2016	33	6	2	23	2	106	13	17	40	36
2015	1	0	0	0	1	221	35	31	128	27

Notes: these underlined values are simulated values.

the ProQuest database is denoted by "X01," and the total retail sales in the U.S. are denoted by "Y01."

In addition, time series data for "X01" to "X70" and "Y01" to "Y08" are needed for considering further disposals because some variables are rare or data are missing. The method of disposal for these situations is the following: (1) the range of time series data from theoretical research is from 1990 to 2017, and retrieval results without data in the literature database are filled by default with zeroes; (2) a small amount of data are available for analysis, such as for variables X06 to X10 and X23 to X56; and (3) data are missing in aspects of retail sales in the U.S. and China from variables Y01 to Y08, which are estimated by the linear trend method at the missing point, as shown in Table 4; additionally, partial simulated values are generated for variables Y01 to Y08, as shown in Table 5. Of course, when determining model variables, we do not consider all variables in various models and select and analyze the main regression model results through repeated tests.

2.3. Descriptive Statistics. Table 6 shows that X06 and X07's Cronbach's alpha value is 0.840, which is greater than 0.5 [17] and indicates consistency for the ProQuest literature database and WoS (Core Collection) regarding the fact that X01 and X02's Cronbach's alpha value is under 0.5, similar to CNKI (all), CNKI (periodical), and CNKI (master's and doctoral dissertation). Consequently, choosing objects from the WoS (Core Collection) and CNKI (periodical) for a theoretical comparison of the U.S. and China is sound and representative. Most variables are not good for data when their std. deviations are greater than their mean. Except for X01–X05, X12, X21, X22, X57, X63, X64, X68, X69, and X70, the others show a Kolmogorov-Smirnov Sig.<0.05 [18, 19] and have an abnormal distribution. Moreover, several methods, including converting to a normal distribution of the data [20], adopting an appropriate regression model and regression standardized residual test [21], and employing nonparametric tests [22, 23], can be used, as discussed in the next section.

As Table 7 demonstrates, data for practical activities, such as relevant retail e-commerce sales and the number of data companies founded, exhibit abnormal distributions. Only variables Y01, Y03, Y08, and Y015's show a Kolmogorov-Smirnov Sig.>0.05, and the others are lower than 0.05. In addition, the other statistics from these variables are similar to those of theoretical research variables and will be discussed regarding validation in further regression.

2.4. Linear Regression Model. Multiple regression analysis is used to determine the relationship between the dependent variables Y and independent variables X; both are time series variables [24]:

Table 6: Summary statistics, Cronbach's alpha values, and test results of normality for the main variables whose measure is theoretical research.

Variable	Mean	S.D.	Cronbach's alpha	N of items	Kolmogorov-Smirnov[a]
X01	1578.61	1999.45	0.123	2	Where X01-X05, X12, X21, X22, X57, X63, X64, X68, X69, X70, Sig.>0.05; others Sig.<0.05.
X02	220.39	171.21			
X03	565.82	546.48			
X04	283.00	282.99	0.877	3	
X05	131.39	172.08			
X06	1.36	2.83	0.842	2	
X07	1.29	3.52			
X08	2.79	7.05			
X09	1.71	4.86	0.831	3	
X10	0.61	1.85			
X11, X13, X15, …, X65, X67, X69			0.889	30	
X12, X14, X16, …, X66, X68, X70			0.889	30	

[a]Lilliefors significance correction.

Table 7: Summary statistics and test results of normality for the variables whose measure is practical activities.

Variable	N	Mean	S.D.	Median	Minimum	Maximum	Kolmogorov-Smirnov[a]		Shapiro-Wilk	
							df	Sig.	df	Sig.
Y01	28	3346.30	980.22	3367.89	1751.27	5076.00	28.00	0.200*	28	0.398
Y02	28	115.42	132.25	64.88	0.03	453.46	28.00	0.01	28	0.001
Y03	28	2.71	2.76	1.92	0.00	8.93	28.00	0.054	28	0.004
Y04	28	50.94	46.12	26.45	2.76	190.73	28.00	0	28	0
Y05	28	10680.61	10786.62	5600.87	830.01	36626.20	28.00	0.001	28	0
Y06	28	855.70	1790.16	5.95	0.00	7175.10	28.00	0	28	0
Y07	28	3.01	5.36	0.10	0.00	19.59	28.00	0	28	0
Y08	28	134.52	70.87	147.90	10.36	242.61	28.00	0.200*	28	0.098
Y09	28	16.96	15.09	10.50	0	51	28.00	0.011	28	0.004
Y10	28	3.43	3.10	2	0	10	28.00	0.002	28	0.007
Y11	28	1.07	1.09	1	0	4	28.00	0	28	0.001
Y12	28	11.39	11.00	6.50	0	37	28.00	0.001	28	0.002
Y13	28	1.11	1.37	1	0	5	28.00	0	28	0
Y14	28	61.21	64.84	40	0	230	28.00	0.018	28	0.001
Y15	28	13.29	10.74	12	0	35	28.00	0.200*	28	0.042
Y16	28	10.36	10.66	6	0	35	28.00	0.008	28	0.002
Y17	28	32.00	37.70	19	0	140	28.00	0	28	0
Y18	28	5.61	9.11	2	0	36	28	0	28	0

[a]Lilliefors significance correction; *This is a lower bound of true significance.

$$y_t = \alpha_0 + \beta_i x_{it} + e_t, \quad (1)$$

where y_t denotes the tth year observation of the dependent variable and x_{it} is a column vector of observations on i independent of the tth year. Four model specification techniques are used to select the variables in a regression model, all possible regression, forward selection, backward elimination, and stepwise regression, showed by Jomnonkwao et al. [24].

2.5. Regression Models of the U.S. Putting Theoretical Research into Practice in E-Commerce

2.5.1. Linear Regression Model for Retail Sales with Theoretical Research Variables. First, an investigation of the promoting effect of theoretical research on retail sales from these normal distribution variables, which include X02, X21, X57, X63, and X69, is conducted by means of running a stepwise regression [25, 26] in SPSS. And it is found the probability of F to enter ≤0.05 [27] of X63, X69, and X21.

Table 8: Model summary for Y01[e].

Model	R	R square	Adjusted R square	Std. error of the estimate	Durbin-Watson
1	0.922[a]	0.849	0.843	387.859	
2	0.939[b]	0.882	0.873	349.711	
3	0.950[c]	0.903	0.891	323.319	
4	0.950[d]	0.903	0.895	317.882	1.604

[a]Predictors: (constant), X63; [b]predictors: (constant), X63, X69; [c]predictors: (constant), X63, X69, X21; [d]predictors: (constant), X69, X21; [e]dependent variable: Y01. X21: searching subject term "Apriori" in the WoS (Core Collection). X63: searching subject term "Mobile Technology & E-Commerce" in the WoS (Core Collection). X69: searching subject term "Quantum Computing" in the WoS (Core Collection). Y01: total retail sales for the U.S.

Table 9: ANOVA for Y01[a].

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	22031078.223	1	22031078.223	146.449	0.000[b]
	Residual	3911305.145	26	150434.813		
	Total	25942383.368	27			
2	Regression	22884937.837	2	11442468.919	93.562	0.000[c]
	Residual	3057445.531	25	122297.821		
	Total	25942383.368	27			
3	Regression	23433533.702	3	7811177.901	74.723	0.000[d]
	Residual	2508849.666	24	104535.403		
	Total	25942383.368	27			
4	Regression	23416156.434	2	11708078.217	115.865	0.000[e]
	Residual	2526226.934	25	101049.077		
	Total	25942383.368	27			

[a]Dependent variable: Y01; [b]predictors: (constant), X63; [c]predictors: (constant), X63, X69; [d]predictors: (constant), X63, X69, X21; [e]predictors: (constant), X69, X21.

Four models have an excellent fit, with all achieving $R^2 > 0.8$ for evaluating the dependent variable Y01 and the three independent variables, as shown in Table 8. In Table 9, the dependent variable and independent variables have a linear relation, where all Sig. of the F statistics [28] are less than 0.01, and they are available for the predictive analysis adopted by models 1, 2, and 4, where the tolerance values are the same in the multiple linear regression model and their VIF values are less than five, as shown in Table 10, which indicates there is no collinearity [29] between the independent variables, which can also be confirmed in Table 11 by the different eigenvalue and its variance proportions for the independent variables. Moreover, the standardized residuals of the regression are normally distributed [30].

Distinctly, theoretical research in the U.S. promotes dramatically practical activities in e-commerce. For example, we can more clearly understand the role of theoretical research in driving the development of retail e-commerce, from the information provided in Tables 12 and 13, in which the literature on "E-Commerce" and "E-Commerce" and "Big data" has positive effects on retail e-commerce sales in practice.

2.5.2. Investigation of the Moderation and Mediation Effects of Variables of Data Companies Founded in the U.S. While running several stepwise regressions for variables of the founded number of data companies and some of the random theoretical research variables, it was found that these variables do not fit very well. For example, by selecting X2, X7, X13, X27, X29, X33, X55, X57, X59, X61, X63, X65, X67, and X69 as independent variables in accordance with the theoretical analysis, where the dependent variables are Y09, Y10, Y11, Y12, and Y13 in sequence, multiple linear regression models are constructed. And the results show that only the independent variable X02 was retained in the regression model along with the dependent variable, such as Y09, Y10, Y11, Y12, or Y13. However, the largest R square in these models is 0.407, which is less than 0.5. Likewise, the results show that only dependent variable Y12 is significantly related to X33 and X55 (model 2 in Table 14) while selecting X7, X13, X27, X29, X33, X55, X57, X59, X61, X63, X65, X67, and X69 as independent variables, and R square is 0.787. In summary, the variables representing the number of data companies founded do not have a good linear relation to the theoretical research variables. Therefore, their moderation and mediation effects [31, 32] will be investigated in the following section.

Here, models of two regression equations are presented, one of which is made up of the independent variables X63 and X69, moderator variable Y13, and dependent variable Y02, and the other model has the interaction term X69Y13 added. It is determined whether the moderator variable has an effect on the relationship between independent variables and dependent variables or not by judging the significance of the R square change (Sig. F change $= 0.032 < 0.05$ in Table 15), which indicates that the data companies founded (research/consulting company) in the U.S. played a moderating role [31, 32] in the theoretical research work of "e-commerce and information technology" in promoting retail e-commerce sales in practice.

Then, we test the mediation effects [31, 32] of the data companies founded variables in the U.S. as an example, such that X02 is the independent variable (shown in Table 12), Y02 is the dependent variable, and one of the data companies founded variables (Y09–Y13) is the mediating variable. The first step is the regression of Y02 on X02, which has a regression standardized coefficient of 0.797 (Sig. $= 0.000 <$ 0.05, $R^2 = 0.635$); the second step is running the linear regression of the independent variable X02 and the dependent variables as one of Y09–Y13, for which all of the regression coefficients are found to be significant, less than 0.05; the last step is building a linear regression of the independent variable X02 and adding one of Y09–Y13, for which the dependent variable is Y02, and all of Y09–Y13 are found to be not significant in this regression model. Hence, a further Sobel test to judge whether variables Y09–Y13 enjoy the mediation effect or not should be performed. The results of the Sobel test shown in Table 16 indicate that the data company variables are not significant regarding their

Table 10: Coefficients for Y01[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	2421.955	105.862		22.878	0.000					
	X63	43.208	3.570	0.922	12.102	0.000	0.922	0.922	0.922	1.000	1.000
2	(Constant)	2196.773	127.959		17.168	0.000					
	X63	32.697	5.118	0.697	6.389	0.000	0.922	0.788	0.439	0.396	2.527
	X69	7.940	3.005	0.288	2.642	0.014	0.830	0.467	0.181	0.396	2.527
3	(Constant)	2129.276	121.916		17.465	0.000					
	X63	5.252	12.881	0.112	0.408	0.687	0.922	0.083	0.026	0.053	18.729
	X69	11.969	3.288	0.435	3.640	0.001	0.830	0.596	0.231	0.283	3.540
	X21	26.882	11.735	0.502	2.291	0.031	0.875	0.424	0.145	0.084	11.918
4	(Constant)	2114.929	114.764		18.428	0.000					
	X69	12.960	2.179	0.471	5.948	0.000	0.830	0.765	0.371	0.622	1.608
	X21	31.332	4.238	0.585	7.393	0.000	0.875	0.828	0.461	0.622	1.608

[a]Dependent variable: Y01, model 3 indicates collinearity.

Table 11: Collinearity diagnostics for Y01[a].

Model	Dimension	Eigenvalue	Condition index	Variance proportions			
				(Constant)	X63	X69	X21
1	1	1.722	1.000	0.14	0.14		
	2	0.278	2.486	0.86	0.86		
2	1	2.648	1.000	0.03	0.02	0.01	
	2	0.282	3.067	0.55	0.28	0.00	
	3	0.070	6.145	0.42	0.70	0.98	
3	1	3.456	1.000	0.02	0.00	0.01	0.00
	2	0.416	2.883	0.33	0.01	0.01	0.03
	3	0.112	5.546	0.52	0.00	0.40	0.08
	4	0.015	15.046	0.14	0.99	0.59	0.89
4	1	2.548	1.000	0.03		0.02	0.04
	2	0.341	2.732	0.34		0.00	0.59
	3	0.111	4.794	0.63		0.97	0.37

[a]Dependent variable: Y01.

Table 12: Coefficients for Y02[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-20.263	25.362		-0.799	0.432					
	X02	0.616	0.091	0.797	6.729	0.000	0.797	0.797	0.797	1.000	1.000
2	(Constant)	-0.255	17.475		-0.015	0.988					
	X02	0.410	0.072	0.530	5.714	0.000	0.797	0.753	0.457	0.742	1.347
	X07	19.772	3.490	0.526	5.665	0.000	0.795	0.750	0.453	0.742	1.347

[a]Dependent variable: Y02, where model 1 R^2 is 0.635 and model 2 is 0.840; low VIF values indicate low collinearity; the standardized residuals are approximately normally distributed. X02: searching subject term "E-Commerce" in the WoS (Core Collection). X07: searching subject term "E-Commerce" and then "Big data" in the WoS (Core Collection). Y02: retail e-commerce sales.

Table 13: Coefficients for Y03[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	0.419	0.204		2.055	0.050					
	X21	0.144	0.009	0.958	16.985	0.000	0.958	0.958	0.958	1.000	1.000
2	(Constant)	0.251	0.158		1.588	0.125					
	X21	0.093	0.013	0.617	7.162	0.000	0.958	0.820	0.304	0.244	4.103
	X22	0.027	0.006	0.392	4.557	0.000	0.929	0.674	0.194	0.244	4.103
3	(Constant)	0.048	0.138		0.347	0.732					
	X21	0.045	0.016	0.295	2.704	0.012	0.958	0.483	0.093	0.098	10.172
	X22	0.024	0.005	0.344	4.877	0.000	0.929	0.706	0.167	0.236	4.240
	X63	0.051	0.013	0.387	3.805	0.001	0.956	0.613	0.130	0.114	8.793

[a]Dependent variable: Y03, where model 1 R^2 is 0.917, model 2 is 0.955, and model 3 is 0.972; low VIF values indicate low collinearity; the standardized residuals are approximately normally distributed. X21: searching subject term "Apriori" in the WoS (Core Collection). X22: searching subject term "Apriori" in CNKI (periodical). X63: searching subject term "Mobile Technology & E-Commerce" in the WoS (Core Collection). Y03: proportion of e-commerce in total retail for U.S.

Table 14: Coefficients for Y12[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	6.201	2.251		2.755	0.011					
	X57	0.702	0.195	0.577	3.604	0.001	0.577	0.577	0.577	1.000	1.000
2	(Constant)	2.223	1.406		1.581	0.126					
	X57	2.041	0.215	1.678	9.498	0.000	0.577	0.885	0.876	0.273	3.667
	X33	-0.261	0.036	-1.290	-7.305	0.000	0.140	-0.825	-0.674	0.273	3.667
3	(Constant)	2.316	1.300		1.782	0.087					
	X57	2.028	0.199	1.667	10.213	0.000	0.577	0.902	0.870	0.272	3.670
	X33	-0.331	0.045	-1.632	-7.401	0.000	0.140	-0.834	-0.631	0.149	6.698
	X55	0.846	0.367	0.402	2.303	0.030	0.230	0.425	0.196	0.238	4.196

[a]Dependent variable: Y12, where model 1 R^2 is 0.333, model 2 is 0.787, and model 3 is 0.826; model 3 indicates collinearity; the standardized residuals are approximately normally distributed. X33: searching subject term "Cloud" and "Big Data" in the WoS (Core Collection). X55: searching subject term "E-commerce & Big Data Analytics" in CNKI (periodical). X57: searching subject term "Online Consumer Behavior" classified by the field of "title" in the WoS (Core Collection). Y12: founded number of data companies in the U.S. classified as industrial application.

mediation effects on the given regression for the fields of theoretical research and practical activities.

2.6. Regression Models of China Putting Theoretical Research into Practice in E-Commerce

2.6.1. Linear Regression Model for Retail Sales with Theoretical Research Variables. The analysis procedure for China is the same as that performed for the U.S. First, the regression is run of the normally distributed variables Y08 and X04, X12, X22, X64, X68, and X70 by means of stepwise regression. And then, the two models are found to have an excellent fit, with all achieving $R^2 > 0.7$, as shown in Table 17. The relationship between X68 and Y08 is positive, but their relationship with X04 is negative. Similarly, China's theoretical research promotes dramatically practical activities in e-commerce. For example, Table 18 shows that the literature on "E-Commerce," "Business Intelligence Analytics," "Mobile Technology and E-Commerce," "Artificial Intelligence and Big Data and E-Commerce," "Quantum Computing," etc., all have positive effects on retail e-commerce sales in practice.

2.6.2. Investigation of the Moderating and Mediating Effects of the Variables of Data Companies Founded in China. While the results of running several stepwise regressions for the variables of the founded number of data companies (Y14, Y15, Y16, Y17, and Y18) and some of the random theoretical research variables (X4, X9, X14, X28, X30, X34, X56, X58, X60, X62, X64, X66, X68, and X70) are the same as those for the U.S., the results indicate that China's theoretical research promotes data companies that are founded in an obvious and direct manner, which can be seen in Table 19. This observation is opposite to that in the U.S. in this regard.

Next, we performed an investigation of the moderation and mediation effects [31, 32] of data companies. First, the

Table 15: Model summary for the moderation effect[d].

Model	R	R square	Adjusted R square	Std. error of the estimate	Change statistics					Durbin-Watson
					R square change	F change	df1	df2	Sig. F change	
1	0.949[a]	0.901	0.897	42.48186	0.901	235.664	1	26	0.000	
2	0.959[b]	0.921	0.914	38.73184	0.020	6.278	1	25	0.019	
3	0.973[c]	0.946	0.934	33.92368	0.026	3.530	3	22	0.032	1.783

[a]Predictors: (constant), X63; [b]predictors: (constant), X63, X69; [c]predictors: (constant), X63, X69, Y13, X69 × Y13, X63 × Y13, and the Sig. of interaction term X69 × Y13 in the regression model is 0.025, less than 0.05; [d]dependent variable: Y02; X63: searching subject term "Mobile Technology & E-Commerce" in the WoS (Core Collection). X69: searching subject term "Quantum Computing" in the WoS (Core Collection). Y13: founded number of data companies in the U.S. classified as research/consulting. Y02: retail e-commerce sales for the U.S.

Table 16: Sobel test for the mediation effects.

Mediation variable																				
Y09				Y10				Y11				Y12				Y13				
Input		Sobel test		Input		Sobel test		Input		Sobel test		Input		Sobel test		Input		Sobel test		
a	0.056	Test statistic	0.726	a	0.011	Test statistic	0.196	a	0.003	Test statistic	0.379	a	0.038	Test statistic	1.039	a	0.004	Test statistic	-1.079	
b	1.001	Std. error	0.077	b	1.304	Std. error	0.073	b	6.309	Std. error	0.050	b	1.906	Std. error	0.070	b	-14.736	Std. error	0.055	
S_a	0.013	P value	0.468	S_a	0.003	P value	0.845	S_a	0.001	P value	0.704	S_a	0.010	P value	0.299	S_a	0.001	P value	0.280	
S_b	1.360			S_b	6.650			S_b	16.496			S_b	1.765			S_b	13.145			

a = raw (unstandardized) regression coefficient for the association between X02 and mediator; s_a = standard error of a; b = raw coefficient for the association between the mediator and the Y02 (when the X02 is also a predictor of the Y02); s_b = standard error of b; calculation for the Sobel test from online, http://quantpsy.org/sobel/sobel.htm. Y09: founded number of data companies in the U.S. (total); Y10: founded number of data companies in the U.S. classified as data/technology; Y11: founded number of data companies in the U.S. classified as business analytics; Y12: founded number of data companies in the U.S. classified as industrial application; Y13: founded number of data companies in the U.S. classified as research/consulting.

Table 17: Coefficients for Y08[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	196.791	9.301		21.159	0.000					
	X04	-0.220	0.023	-0.879	-9.383	0.000	-0.879	-0.879	-0.879	1.000	1.000
2	(Constant)	199.906	8.734		22.887	0.000					
	X04	-0.317	0.048	-1.268	-6.631	0.000	-0.879	-0.798	-0.576	0.206	4.844
	X68	4.175	1.828	0.437	2.284	0.031	-0.693	0.416	0.198	0.206	4.844

[a]Dependent variable: Y08, where model 1 R^2 is 0.772 and model 2 is 0.811; low VIF values indicate low collinearity; the standardized residuals are approximately normally distributed. X04: searching subject term "E-Commerce" in CNKI (periodical); X68: searching subject term "Artificial Intelligence & Big Data & E-Commerce" in CNKI (periodical); Y08: growth rate of retail e-commerce for China.

moderation of data company variables is judged in the two regression equations, one of which is composed of the dependent variables X04, X12, and X68; moderator variable Y15; and independent variable Y06, and the other equation has the added interaction term, X04Y15, X12Y15, or X68Y15. Then, the decision regarding whether the moderator variable has an effect on the relationship between the independent variables and dependent variables or not is made according to the significance of the R square change. The X04Y15 regression model indicates that the R square change is valid (Sig. F change = 0.000 < 0.05, regression coefficient Sig. = 0.000 < 0.05), the X12Y15 is Sig. F change = 0.032 < 0.05 and regression coefficient Sig. = 0.010 < 0.05, and the X68Y15 is Sig. F Change = 0.000 < 0.05 and regression coefficient Sig. = 0.000 < 0.05, which indicates that the data companies founded variable (data/technology) in China has a moderating effect on theoretical research work promoting retail e-commerce sales in practice. We also test the mediation effects of the variables of data companies founded in China as an example, using X04 as the independent variable, Y06 as the dependent variable, and one of the data companies founded variables (Y14–Y18) as the mediating variable. The first step is the regression of Y06 on X04, which has a standardized regression coefficient of 0.842

Table 18: Regression models for China's theoretical research and practical activities in e-commerce.

Dependent variable	Independent variable					Statistics
	X04	X12	X64	X68	X70	
Y05	1.153*** $t = 11.358$ Sig. = 0.000				-0.221* $t = -2.1788$ Sig. = 0.039	$R^2 = 0.936$, both VIF = 4.035, $F = 183.264$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y06	0.451* $t = -2.670$ Sig. = 0.013	-0.343*** $t = -3.647$ Sig. = 0.001		0.695*** $t = 4.480$ Sig. = 0.000		$R^2 = 0.881$, VIF (X04) = 5.763, VIF (X12) = 1.787, VIF (X68) = 4.851, $F = 59.268$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y07	1.497*** $t = 8.762$ Sig. = 0.000		-0.669*** $t = -3.912$ Sig. = 0.001			$R^2 = 0.872$, both VIF = 5.686, $F = 84.834$ (Sig.<0.001), the standardized residuals are approximately normally distributed

*Sig.<0.05, **Sig.<0.01, and ***Sig.<0.001; low VIF values indicate low collinearity; all coefficients are standardized coefficients. X04: searching subject term "E-Commerce" in CNKI (periodical). X12: searching subject term "Business Intelligence Analytics" in CNKI (periodical). X64: searching subject term "Mobile Technology & E-Commerce" in CNKI (periodical). X68: searching subject term "Artificial Intelligence & Big Data & E-Commerce" in CNKI (periodical). X70: searching subject term "Quantum Computing" in CNKI (periodical). Y05: total retail sales for China. Y06: retail e-commerce sales for China. Y07: proportion of e-commerce in total retail for China.

Table 19: Regression models for China's theoretical research and data companies founded.

Dependent variable	Independent variable							Statistics
	X04	X09	X28	X60	X64	X66	X68	
Y14					0.418*** $t = 4.522$ Sig. = 0.000	0.596*** $t = 6.438$ Sig. = 0.000		$R^2 = 0.911$, both VIF = 2.413, $F = 128.417$ (sig.<0.001), the standardized residuals are approximately normally distributed
Y15	1.136*** $t = 12.353$ Sig. = 0.000			-0.480*** $t = - -5.223$ Sig. = 0.000				$R^2 = 0.865$, both VIF = 1.566, $F = 80.108$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y16					0.530*** $t = 4.435$ Sig. = 0.000	0.452*** $t = 3.780$ Sig. = 0.001		$R^2 = 0.852$, both VIF = 2.413, $F = 71.915$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y17		-0.210* $t = -2.368$ Sig. = 0.026			0.351** $t = 3.203$ Sig. = 0.004	0.755*** $t = 6.194$ Sig. = 0.000		$R^2 = 0.881$, VIF (X09) = 1.590, VIF (X64) = 2.413, VIF (X66) = 2.992, $F = 59.053$ (Sig.<0.001), the standardized residuals are approximately normally distributed
Y18			1.069*** $t = 12.345$ Sig. = 0.000		0.418*** $t = 6.796$ Sig. = 0.000		-0.469*** $t = -4.863$ Sig. = 0.000	$R^2 = 0.958$, VIF (X28) = 4.321, VIF (X64) = 2.180, VIF (X68) = 5.353, $F = 183.898$ (Sig.<0.001), the standardized residuals are approximately normally distributed

*Sig.<0.05, **Sig.<0.01, and ***Sig.<0.001; low VIF values indicate low collinearity; all coefficients are standardized coefficients; X04: searching subject term "E-Commerce" in CNKI (periodical); X09: searching subject term "E-Commerce" and then "Big data" in CNKI (periodical); X28: searching subject term "Machine Learning" and "Big Data" in CNKI (periodical); X60: searching subject term "Online Consumer Behavior & Big Data" classified by the field of "topic" in CNKI (periodical); X64: searching subject term "Mobile Technology & E-Commerce" in CNKI (periodical); X66: searching subject term "Cloud Computing & E-Commerce" in CNKI (periodical); X68: searching subject term "Artificial Intelligence & Big Data & E-Commerce" in CNKI (periodical); Y14: founded number of data companies in China (total); Y15: founded number of data companies in China classified as data/technology; Y16: founded number of data companies in China classified as business analytics; Y17: founded number of data companies in China classified as industrial application; Y18: founded number of data companies in China classified as research/consulting.

Table 20: Coefficients for Y06[a].

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	Correlation			Collinearity statistics	
		B	Std. error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-89.898	208.579		-0.431	0.670					
	Y18	168.642	19.750	0.859	8.539	0.000	0.859	0.859	0.859	1.000	1.000
2	(Constant)	-440.918	243.502		-1.811	0.082					
	Y18	100.833	34.095	0.513	2.957	0.007	0.859	0.509	0.274	0.286	3.500
	X04	2.584	1.098	0.408	2.353	0.027	0.842	0.426	0.218	0.286	3.500

[a]Dependent variable: Y06, where model 1 R^2 is 0.737 and model 2 is 0.785; low VIF values indicate low collinearity; the standardized residuals are approximately normally distributed. X04: searching subject term "E-Commerce" in CNKI (periodical). Y18: founded number of data companies in China classified as research/consulting. Y06: retail e-commerce sales for China.

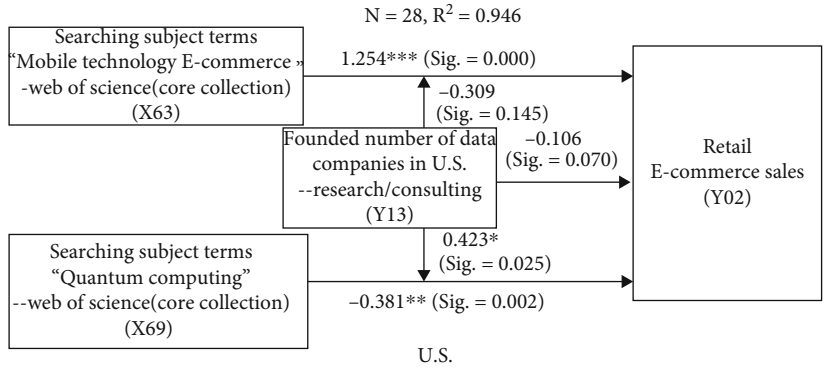


Figure 1: Moderating effect of data companies founded in the U.S.

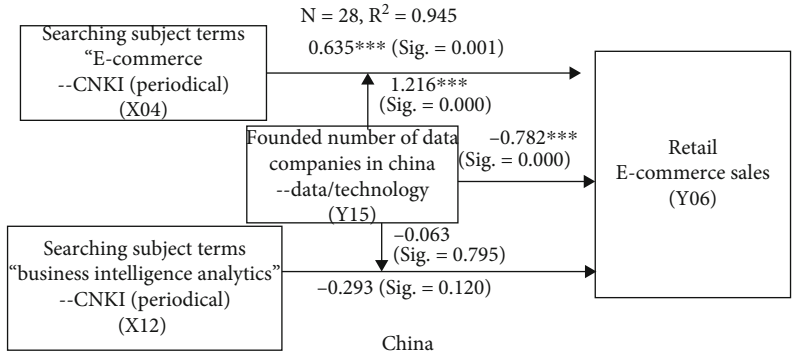


Figure 2: Moderating effect of data companies founded in China.

(Sig. = 0.000 < 0.05, R^2 = 0.710); the second step is running a linear regression of the independent variable X04 and dependent variables from one of Y14–Y18, which shows that all of the regression coefficients are significant, less than 0.05. The last step is building a linear regression of the independent variable X04 and adding one of Y14–Y18, for which the dependent variable is Y06, and it is found that only Y18 (Sig. = 0.007 < 0.05) and X04 (Sig. = 0.027 < 0.05) are simultaneously significant in this regression model (shown in Table 20). As a consequence, data companies in China play a mediating role in putting these fields of theoretical research into practice in e-commerce.

2.7. Comparison of the Moderating and Mediating Effects of Data Companies Founded between the U.S. and China. In Figures 1 and 2, we can see that both the U.S. and China show significance in moderating the relationship between theoretical research and practice in e-commerce using BDA, as tested by the moderate variable "founded number of data companies." In the U.S., the variable "founded number of data companies" has a moderating effect on the model of the correlation of "searching subject terms 'Quantum Computing'—WoS (Core Collection) (X69)" and "retail e-commerce sales (Y02)," such that this relationship is negative (X69: -0.381 (Sig. = 0.002 < 0.05); Y13: 0.423

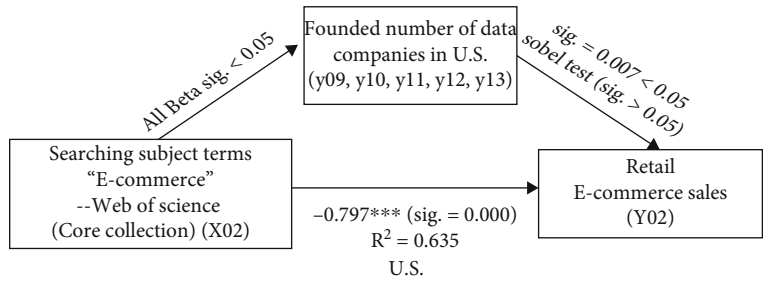


Figure 3: Mediating effect of data companies founded in the U.S.

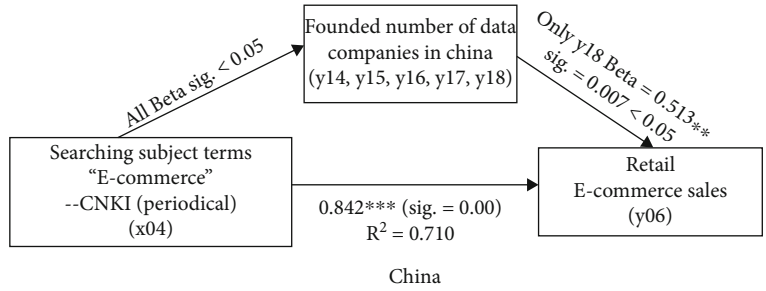


Figure 4: Mediating effect of data companies founded in China.

Table 21: Lag order selection criteria.

Variable	Lag	Log*L*	LR	FPE	AIC	SC	HQ	Lag order selection	Observed object
X02	0	-162.121	NA	27223.100	13.050	13.098	13.063		U.S.
	1	-139.237	42.107*	4728.674*	11.299*	11.396*	11.326*	○	
	2	-138.811	0.749	4955.199	11.345	11.491	11.385		
	3	-138.625	0.313	5297.007	11.410	11.605	11.464		
X69	0	-121.768	NA	1078.827	9.821	9.870	9.835		
	1	-108.126	25.101	392.517	8.810	8.908	8.837		
	2	-103.429	8.267*	292.255*	8.514*	8.661*	8.555*	○	
	3	-103.429	0.000	317.104	8.594	8.789	8.648		
X04	0	-175.910	NA	82037.320	14.153	14.202	14.166		China
	1	-130.324	83.878*	2317.735*	10.586*	10.683*	10.613*	○	
	2	-130.303	0.036	2508.760	10.664	10.811	10.705		
	3	-130.278	0.043	2716.554	10.742	10.937	10.796		
X70	0	-83.390	NA	50.069	6.751	6.800	6.765		
	1	-73.001	19.117*	23.630	6.000	6.098	6.027		
	2	-71.283	3.022	22.330*	5.943*	6.089*	5.983*	○	
	3	-70.835	0.753	23.376	5.987	6.182	6.041		

∗ indicates lag order selected by the criterion. LR: sequential modified LR test statistic (each test at 5% level). FPE: final prediction error; AIC: Akaike information criterion; SC: Schwarz information criterion; HQ: Hannan-Quinn information criterion. Sample: 1990-2017. X02: searching subject term "E-Commerce" in the WoS (Core Collection). X04: searching subject term "E-Commerce" in CNKI (periodical). X69: searching subject term "Quantum Computing" in the WoS (Core Collection). X70: searching subject term "Quantum Computing" in CNKI (periodical).

(Sig. = 0.025 < 0.05)), which means that it does not moderate the correlation of "searching subject terms 'Mobile Technology and E-Commerce'—WoS (Core Collection) (X63)" and "retail e-commerce sales (Y02)" (Y13: 0.309 (Sig. = 0.145 > 0.05)). This finding is similar to that in China, where the variable "founded number of data companies" has a moderating effect on the model of the correlation of "searching subject terms 'E-Commerce'—CNKI (periodical) (X04)" and "retail e-commerce sales (Y06)," such that if the relationship is positive (X04: 0.635 (Sig. = 0.001 < 0.05); Y15: 1.216 (Sig. = 0.000 < 0.05)), it has a direct significant negative correlation in relation to "retail e-commerce sales

Table 22: Comparison of linear regression models involving lag variables.

Dependent variable	Independent variable			
	X02	X02 (-1)	X69	X69 (-2)
Y02	0.616***	0.624***	2.410***	2.376***
	S.E. = 0.091	S.E. = 0.099	S.E. = 0.554	S.E. = 0.614
	t-statistic = 6.729	t-statistic = 6.282	t-statistic = 4.348	t-statistic = 3.870
	R-squared = 0.635	R-squared = 0.612	R-squared = 0.421	R-squared = 0.384
	F-statistic = 45.277	F-statistic = 39.462	F-statistic = 18.908	F-statistic = 14.981
	X04	X04 (-1)	X70	X70 (-2)
Y06	5.328***	5.868***	170.442***	150.349**
	S.E. = 0.669	S.E. = 0.798	S.E. = 35.220	S.E. = 53.494
	t-statistic = 7.969	t-statistic = 7.357	t-statistic = 4.925	t-statistic = 2.811
	R-squared = 0.710	R-squared = 0.684	R-squared = 0.483	R-squared = 0.248
	F-statistic = 63.507	F-statistic = 54.120	F-statistic = 24.251	F-statistic = 7.899

**Sig.<0.01; ***Sig.<0.001. X02: searching subject term "E-Commerce" in the WoS (Core Collection). X04: searching subject term "E-Commerce" in CNKI (periodical). X69: searching subject term "Quantum Computing" in the WoS (Core Collection). X70: searching subject term "Quantum Computing" in CNKI (periodical). Y02: retail e-commerce sales for U.S. Y06: retail e-commerce sales for China.

Figure 5: Trend of big data analytics in e-commerce. Website: https://www.amazon.com/Magformers-Magnetic-Building-Educational-Construction/dp/B06XJLGWST?ref_=Oct_DLandingS_PC_NA_NA#customerReviews.

(Y06)" (Y15: -0.782 (Sig. = 0.000 < 0.05)). However, it does not work in the correlation of "searching subject terms 'Business Intelligence Analytics'—CNKI (periodical) (X12)" (X12: -0.293 (Sig. = 0.120 > 0.05)) and "retail e-commerce sales (Y02)" (Y15: -0.063 (Sig. = 0.795 > 0.05)).

The model presented in Figure 3 assumes a three-variable system, which has a direct and significant relationship between "retail e-commerce sales (Y02)" and "searching subject terms 'E-Commerce'—WoS (Core Collection) (X02)," and the mediator variable "founded number of data companies in the U.S." is introduced to the model. However, this path between Y02 and X02 becomes nonsignificant because for "the number of existing data companies in the United States," it plays an important role in promoting the theoretical research of e-commerce in practice. However, data companies in China shown in Figure 4 are limited to

these types of "research/consulting" data companies and have a vivid mediating effect on the relationship between theoretical research works of BDA in e-commerce and practical producing activities in e-commerce.

2.8. Lag Consideration in Evaluating Theoretical Research Response to Practical Application in E-Commerce. In general, this theoretical research puts into practice needs a certain lag to accomplish the task. Here, a selection of the variables X02 and X69 for the U.S. and X04 and X70 for China is made to test the linear relation between their lag and the retail e-commerce sales. First, the goal is to determine the lag order among the independent variables selected from Table 21, which shows the six criteria [33], and the results of the lag order selection for X02 and X69 and X04 and X70 are a lag order of one and two and one and two, respectively. Next, we construct linear regression models involving the lag variables to determine whether the involved lag variables in the regression models fit better or not. The answer is certainly not. No matter whether the U.S. or China is investigated, R square degrades, as shown in Table 22. These findings indicate that we should not consider the effects of lag on evaluating the theoretical research response to practical applications in e-commerce, which was also demonstrated in previous sections as observed by the nonlag variables in the empirical studies.

3. Conclusions

The rapid growth of e-commerce has benefited not only the evolution of data science over the past two decades but also the boom of big data from various sources. This is what makes China and the United States the largest e-commerce markets and why China accounts for more than the United States in e-commerce sales. Ultimately, we can determine the reasons leading to the difference between the U.S. and China regarding this point. One of the reasons is the institutional differences and commercial value, which makes Chinese society's perception of BDA in e-commerce more acceptable than that of the United States. Another reason involves the theoretical research works on BDA in e-commerce in China, which have attracted slightly more extensive attention than that observed in the U.S. and involved a comparison of literature databases, indicating that proof of a significant relationship between theoretical research and practical activities in BDA in e-commerce could be attained. In addition, in the United States, with regard to the relationship of putting theoretical research into practice, the variables of the data company show moderate but no mediating effect. However, in China, the mediating effects of this relationship have been explained. These results help clarify doubts regarding the development of China e-commerce, which even exceeds that of the U.S. today, in view of the theoretical and practical comparison of BDA in e-commerce between them.

3.1. Avenues for Future Research and Practice. Regardless of whether the U.S. or China is considered, the theoretical research work is deeply impressing and has propelled practical application of BDA in e-commerce. However, big data hubris and algorithm dynamics issues may contribute to analysis mistakes [7] because of human subjective prejudice, technological objective limitations, and the need to enhance artificial intelligence by processing data more efficiently for e-commerce transactions. We expect that e-commerce activities concerning the seller, buyer, platform provider, etc., would entail self-learning actively through their own generated data, and then, extraction by others (such as the buyer and platform provider) of the critical information would serve as a combination of commodities for improving the quality of sales and service, particularly for increasing the transparency and credibility of goods to attract purchases. For example, currently, a product from Amazon online is mostly displayed with its price and functions; however, it is anticipated that product information, manufacturer information, seller information, customer information, and even extra payment information or more will be shown in the future, as shown in Figure 5, such that the desire to buy and recommend the right goods for purchasers can be reinforced. Therefore, there are three future research orientations in e-commerce using BDA. First, data originating from e-commerce activities will be considered valuable and tradable resources after being processed by BDA, and either the seller, buyer, or platform provider can enact pricing dynamically with his or her data for sale. In addition, the data trading market and its pricing mechanisms in e-commerce will be researched and widely put into use. Second, a new rule for dynamic pricing for each customer developed by applying BDA in e-commerce can be envisioned, such that a product would sell for a different price on a per-customer basis, enlisting every seller, buyer, and platform provider to accomplish each expectation or revenue maximization. Third, puzzling relationships among purchasing behaviors and consumer habits [34], consumer habits and personalities [35, 36], consumer personalities and the growth environment [37, 38] can be unraveled by using BDA for deep learning in e-commerce trading.

In addition, it is expected that the mixture of large data resources and new technologies will challenge many existing e-commerce problems and find out a better solution. A series of new issues should be focused on, such as quantum computing in e-commerce [39], in which theoretical research works serve to observably promote retail sales, both in the U.S. and China (as seen in Tables 9 and 17). Ronald [40] considers the potential impact that the nascent technology of quantum computing may have on e-commerce, more specifically, designing "encrypt" information in such a way to ensure that an e-commerce trade is safe, offering significant speed-ups for faster search and optimization in the big data age, and implementing the quantum cheque transaction in a quantum-networked banking system [41]. As a BDA concept, quantum machine learning could enable machine learning that is faster than that of classical computers for calculating and analyzing e-commerce activities in the big data age [42]. In short, quantum computing in e-commerce is a crucial theoretical research topic and has practical application both for the U.S. and China at present and in the future.

In future applications, we should encourage data companies to devote efforts to big data business issues required for e-commerce because data companies play a moderation or mediation role in putting theoretical research into practice in e-commerce.

Conflicts of Interest

The author declares that they have no conflicts of interest.

Acknowledgments

This work was supported by key projects of the National Social Science Foundation (No. 19AGL017), Humanities and Social Science Research Project of the Ministry of Education (No. 18YJAZH153), Natural Science Foundation of Fujian (No. 2018J01648), and Development Fund of Scientific Research from Fujian University of Technology (No. GY-S18109), all received from the Chinese government.

References

[1] C. L. Philip Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: a survey on big data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[2] S. Sakr, A. Liu, D. M. Batista, and M. Alomari, "A survey of large scale data management approaches in cloud environments," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 311–336, 2011.

[3] W. A. Günther, M. H. Rezazade Mehrizi, M. Huysman, and F. Feldberg, "Debating big data: a literature review on realizing value from big data," *Journal of Strategic Information Systems*, vol. 26, no. 3, pp. 191–209, 2017.

[4] Z. Wang and Q. Yu, "Privacy trust crisis of personal data in China in the era of big data: the survey and countermeasures," *Computer Law & Security Review*, vol. 31, no. 6, pp. 782–792, 2015.

[5] P. Ohm, "The underwhelming benefits of big data," *University of Pennsylvania Law Review Online*, vol. 161, no. 1, pp. 339–346, 2013.

[6] S. Marcel, "Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health," *The Journal of Infectious Diseases*, vol. 214, no. 4, pp. 399–403, 2016.

[7] D. Lazer, R. Kennedy, G. King, and A. Vespignani, "The parable of Google flu: traps in big data analysis," *Science*, vol. 343, no. 6176, pp. 1203–1205, 2014.

[8] R. V. Hal, "Big data: new tricks for econometrics," *Journal of Economic Perspectives*, vol. 28, no. 2, pp. 3–28, 2014.

[9] O. Ahmed, B. Fatima-Zahra, A. L. Ayoub, and B. Samir, "Big data technologies: a survey," *Journal of King Saud University—Computer and Information Sciences*, vol. 30, no. 4, pp. 431–448, 2018.

[10] A. Deepak, K. Niraj, and P. K. John, "Understanding big data analytics capabilities in supply chain management: unravelling the issues, challenges and implications for practice," *Transportation Research Part E: Logistics and Transportation Review*, vol. 114, pp. 416–436, 2017.

[11] S. Akter and S. F. Wamba, "Big data analytics in e-commerce: a systematic review and agenda for future research," *Electronic Markets*, vol. 26, no. 2, pp. 173–194, 2016.

[12] J. Manyika, M. Chui, B. Brown et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011, June 2018, http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation.

[13] A. Taylor, *The 17 Biggest Data Breaches of the 21st Century*, CSO, 2018, July 2018, https://www.csoonline.com/article/2130877/data-breach/the-biggest-data-breaches-of-the-21st-century.html.

[14] Q. Bin, S. J. Chen, and S. Q. Sun, "Cultural differences in e-commerce," *Journal of Global Information Management*, vol. 11, no. 2, pp. 48–55, 2003.

[15] M. B. Schmidt, A. C. Johnston, K. P. Arnett, J. Q. Chen, and S. Li, "A cross-cultural comparison of U.S. and Chinese computer security awareness," *Journal of Global Information Management*, vol. 16, no. 2, pp. 91–103, 2008.

[16] H. Dai and P. C. Palvi, "Mobile commerce adoption in China and the United States," *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, vol. 40, no. 4, pp. 43–61, 2009.

[17] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory (McGraw-Hill Series in Psychology)*, McGraw-Hill, New York, 1994.

[18] Y. H. Chan, "Biostatistics 102: quantitative data-parametric & non-parametric tests," *Singapore Medical Journal*, vol. 44, no. 8, pp. 391–396, 2003.

[19] A. Ghasemi and S. Zahediasl, "Normality tests for statistical analysis: a guide for non-statisticians," *International Journal of Endocrinology and Metabolism*, vol. 10, no. 2, pp. 486–489, 2012.

[20] D. G. Goring and V. I. Nikora, "Despiking acoustic doppler velocimeter data," *Journal of Hydraulic Engineering*, vol. 128, no. 1, pp. 117–126, 2002.

[21] R. L. Brown, J. Durbin, and J. M. Evans, "Techniques for testing the constancy of regression relationships over time," *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 37, no. 2, pp. 149–163, 1975.

[22] B. M. Henry, "Nonparametric tests against trend," *Econometrica*, vol. 13, no. 3, pp. 245–259, 1945.

[23] A. N. Pettitt, "A non-parametric approach to the change-point problem," *Journal of the Royal Statistical Society Series C (Applied Statistics)*, vol. 28, no. 2, pp. 126–135, 1979.

[24] S. Jomnonkwao, S. Uttra, and V. Ratanavaraha, "Forecasting road traffic deaths in Thailand: applications of time-series, curve estimation, multiple linear regression, and path analysis models," *Sustainability*, vol. 12, no. 1, p. 395, 2020.

[25] R. B. Bendel and A. A. Afifi, "Comparison of stopping rules in forward "stepwise" regression," *Journal of the American Statistical Association*, vol. 72, no. 357, pp. 46–53, 1977.

[26] L. Wilkinson, "Tests of significance in stepwise regression," *Psychological Bulletin*, vol. 86, no. 1, pp. 168–174, 1979.

[27] C. Kontoghiorghes, "Reconceptualizing the learning transfer conceptual framework: empirical validation of a new systemic model," *International Journal of Training and Development*, vol. 8, no. 3, pp. 210–221, 2004.

[28] P. T. Pope and J. T. Webster, "The use of an F-statistic in stepwise regression procedures," *Technometrics*, vol. 14, no. 2, pp. 327–340, 1972.

[29] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York, 2005.

[30] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review/Revue Internationale de Statistique*, vol. 55, no. 2, pp. 163–172, 1987.

[31] A. J. Fairchild and D. P. MacKinnon, "A general model for testing mediation and moderation effects," *Prevention Science*, vol. 10, no. 2, pp. 87–99, 2009.

[32] J. R. Edwards and L. S. Lambert, "Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis," *Psychological Methods*, vol. 12, no. 1, pp. 1–22, 2007.

[33] V. Ivanov and L. Kilian, "A practitioner's guide to lag order selection for VAR impulse response analysis," *Studies in Nonlinear Dynamics & Econometrics*, vol. 9, no. 1, 2005.

[34] M. F. Ji and W. Wood, "Purchase and consumption habits: not necessarily what you intend," *Journal of Consumer Psychology*, vol. 17, no. 4, pp. 261–276, 2007.

[35] W. Wood, "Habit in personality and social psychology," *Personality and Social Psychology Review*, vol. 21, no. 4, pp. 389–403, 2017.

[36] H. K. Harold, "Personality and consumer behavior: a review," *Journal of Marketing Research*, vol. 8, no. 4, pp. 409–418, 1971.

[37] D. Scott and F. K. Willits, "Environmental attitudes and behavior: a Pennsylvania survey," *Environment and Behavior*, vol. 26, no. 2, pp. 239–260, 1994.

[38] S. Bhate, "One world, one environment, one vision: are we close to achieving this? An exploratory study of consumer environmental behaviour across three countries," *Journal of Consumer Behaviour: An International Research Review*, vol. 2, no. 2, pp. 169–184, 2002.

[39] K. Thapliyal and A. Pathak, "Quantum e-commerce: a comparative study of possible protocols for online shopping and other tasks related to e-commerce," 2018, July 2018, https://arxiv.org/pdf/1807.08199.pdf.

[40] D. W. Ronald, "The potential impact of quantum computers on society," *Ethics and Information Technology*, vol. 19, no. 4, pp. 271–276, 2017.

[41] B. K. Behera, A. Banerjee, and P. K. Panigrahi, "Experimental realization of quantum cheque using a five-qubit quantum computer," *Quantum Information Processing*, vol. 16, no. 12, p. 312, 2017.

[42] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5754322, 9 pages
https://doi.org/10.1155/2021/5754322

Research Article

Artificial Intelligence- (AI-) Enabled Internet of Things (IoT) for Secure Big Data Processing in Multihoming Networks

Geetanjali Rathee,[1] **Adel Khelifi**,[2] **and Razi Iqbal**[3]

[1]*Department of Computer Science and Engineering, Netaji Subhas University of Technology, Dwarka, Sector-3, Delhi, India*
[2]*Department of Computer Science and Information Technology, College of Engineering, Abu Dhabi University, UAE*
[3]*Department of Computer Information Systems, University of the Fraser Valley, Canada*

Correspondence should be addressed to Razi Iqbal; razi.iqbal@ieee.org

Received 11 June 2021; Revised 11 July 2021; Accepted 3 August 2021; Published 15 August 2021

Academic Editor: Daniel G. Reina

The automated techniques enabled with Artificial Neural Networks (ANN), Internet of Things (IoT), and cloud-based services affect the real-time analysis and processing of information in a variety of applications. In addition, multihoming is a type of network that combines various types of networks into a single environment while managing a huge amount of data. Nowadays, the big data processing and monitoring in multihoming networks provide less attention while reducing the security risk and efficiency during processing or monitoring the information. The use of AI-based systems in multihoming big data with IoT- and AI-integrated systems may benefit in various aspects. Although multihoming security issues and their analysis have been well studied by various scientists and researchers; however, not much attention is paid towards big data security processing in multihoming especially using automated techniques and systems. The aim of this paper is to propose an IoT-based artificial network to process and compute big data processing by ensuring a secure communication multihoming network using the Bayesian Rule (BR) and Levenberg-Marquardt (LM) algorithms. Further, the efficiency and effect on multihoming information processing using an AI-assisted mechanism are experimented over various parameters such as classification accuracy, classification time, specificity, sensitivity, ROC, and *F*-measure.

1. Introduction

A surge in utilization of smart systems has significantly enhanced efficient processing, reliable communication, and secure transmissions via wireless systems. However, augmentation of data may still encounter various computational and communicational risks in the network. In order to perform an efficient and smooth processing of huge records, a big data term came into existence. Big data is defined as the huge collection of records or information in volume that is exponentially growing with the time [1]. Hence, the traditional data management techniques are not able to perform efficiently. Big data along with certain platforms such as Hadoop and cloud servers may organize and manage the online processing or transmission of records; however, the collection of information from various networks may further lead to various complexity and security risks [2].

The involvement of multiple networks while processing or managing the enormous records introduces a new term known as multihoming networks [3]. It is defined as the involvement of various types of networks while clustering the records of information at a single place [4]. The multihoming is considered an emerging mechanism for clustering multiple records in a network. In addition, the processing and management of big data may further introduce complexity, processing, and security of networks and records while processing at a single place [5].

Numbers of smart-based big data schemes for managing or processing the large-size datasets have been proposed by a number of authors/scientists. Cloud-based Internet of Things (IoT) and Artificial Neural Network (ANN) schemes have been used in big data, network clustering, and multihoming schemes for an efficient and automated control systems [6–9]. Furthermore, a number of schemes in multihoming

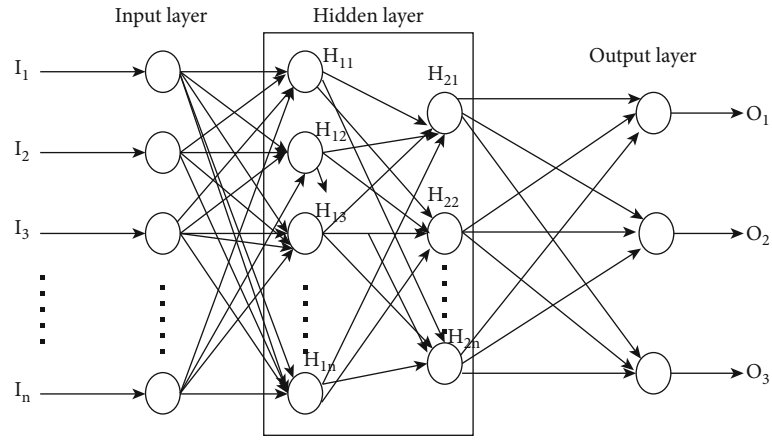


Figure 1: Multilayered perceptron architecture network (this figure is reproduced from Rathee et al. [27]).

and their analysis are proposed by various studies. However, not much attention is given to the automated multihoming schemes for managing, processing, and securing the big data information. In addition, the use of automated and artificial intelligence- (AI-) based smart systems in multihoming networks for securing and processing the information may reduce various security and management risks by enhancing the large-size data distributions, multiple network clustering, and data processing and managing [10]. Further, the AI-enabled proposed mechanism in multihoming for managing, securing, and processing big data provides vast implications for business, research, and future activities [11].

Moreover, the IoT- or ANN-based mechanisms benefitted for processing and monitoring among multihomed sites as depicted in Figure 1. Figure 1 consists of three different layers: input layer, hidden layer, and output layer. The huge amount of data processed by various networks is input in ANN where the processing or analysis of input data is done by the hidden layer. Further, the type of information either trusted or malicious generated by various networks is displayed by the output layer. Furthermore, the ANN-based automated system provides efficient processing, collection, and analysis of huge information while clustering the multiple networks. These techniques may further benefit for preventing, securing, managing, and processing of massive information while ensuring the security in networks [12–14].

1.1. Motivation and Contribution. Numbers of automated schemes have been proposed by various researchers in several applications such as healthcare, IIoT (Industrial Internet of Things), big data, multihoming, and vehicular transportation systems [15]. However, the integration of automated schemes in multihoming networks for managing and processing big data is considered one of the significant research interests. The integration of intelligent techniques while analysing the network algorithms, security, and clustering of multiple networks having various protocols, configurations, and attributes may benefit in a number of techniques. The automated systems may further benefit to manage and store the huge amount of information from heterogenous networks in an efficient way.

The aim of this paper is to propose an efficient and automated AI-based scheme for monitoring and processing various activities, including risk monitoring, processing, and management of big data in multihoming networks. The proposed mechanism is analysed over a simulated synthesized dataset over various processing metrics and parameters such as classification accuracy, classification time, specificity, sensitivity, ROC, and *F*-measure.

The remaining structure of the paper is organized as follows. Section 2 deliberates the number of approaches proposed by various authors/scientists. In addition, the ANN-based proposed phenomenon is detailed in Section 3. Further, Section 4 represents the discussion of results and analysis of the proposed phenomenon against existing mechanisms over various processing metrics. Finally, Section 5 concludes the paper along with future direction.

2. Related Work

The number of schemes proposed by various researchers/scientists is discussed in this section where the technique along with their performance analysis is defined in Table 1. de Santerre et al. [16] have presented an enhanced routing mechanism in multihomed IPv6 terminal sites while dealing with ingress filtering policies. The authors have described a new mechanism for choosing the default route through the packet of the source address. The proposed mechanism resolved the ingress filtering issue without implying Internet service providers. The authors have showed the easy deployment and requirement of changes in terminal nodes during communication. Wang et al. [17] have presented a learning scheme called local mobility anchor by initiating the learning procedures of IP addresses. The proposed mechanism forwarded various interfaces such as active, pending, and home network prefixes at a binding cache entry with learned IP addresses. In order to achieve the IP address, extraction and minimum required messages are detailed and defined by elaborating various charts. Bi et al. [18] have summarized the IPv4 site multihoming limitations and practices by

Table 1: Approaches/schemes proposed by various researchers.

Authors	Technique	Analysis
de Santerre et al. [16]	Enhanced routing mechanism in multihoming	The authors have described a new mechanism for choosing the default route through the packet of the source address
Wang et al. [17]	Learning scheme called local mobility anchor	The proposed mechanism forwarded various interfaces such as active, pending, and home network prefixes at a binding cache entry with learned IP addresses
Bi et al. [18]	IPv4 site multihoming limitations and practices	The authors have discussed various challenges and opportunities to bring up the security and mobility issues in multihoming environment
He et al. [19]	Comprehensive taxonomy of threats	The authors have proposed various criteria for evaluating the data analysis and collection performance
Li et al. [20]	Online IoT security monitoring scheme	An accurate data structural model is proposed to capture the behaviours of smart devices. Further, a hypothesis testing scheme is quantified to monitor the uncertain tasks by resolving the scalability issues
Lin et al. [21]	Network security-related data	The authors have provided the objectives and requirements of information collection with a taxonomy of various data gathering techniques
Wu et al. [22]	Analysis-based architecture	The authors have proposed an authentication mechanism for managing the clusters and enabling the data analysis using a colony optimization scheme
Zhou et al. [23]	Differently private scheme	In order to guarantee the trusted computing, a trust-based mechanism is proposed to evaluate the end user's reliability
Han et al. [24]	Agile confidential transmission strategy	The authors have combined the opportunistic beamforming and driven cluster schemes for managing the huge data collection from various base stations

surveying the current IPv6 multihoming solutions. The authors have discussed various challenges and opportunities to bring up the security and mobility issues in multihoming environment.

He et al. [19] have presented a comprehensive taxonomy of threats according to long-term evolution and advanced network structure. The authors have proposed various criteria for evaluating the data analysis and collection performance. All the traditional schemes and methods have discussed and analysed evaluation criteria by presenting various research and open issues for simulating the schemes. In addition, the authors have proposed a security measurement for analysis and collection of information in long-term evolution and advanced networks. Li et al. [20] have proposed an online IoT security monitoring scheme for distributed networks by selecting an advanced point influential operational abstract. An accurate data structural model is proposed to capture the behaviours of smart devices. Further, a hypothesis testing scheme is quantified to monitor the uncertain tasks by resolving the scalability issues. The authors have committed the cyberthreats to an IoT system for sensing the testbed using various strengths and patterns. The proposed algorithms claimed the efficient monitoring and detection of cyberthreats in IoT-based systems. Lin et al. [21] have introduced the network security-related data having different characteristics and definitions. The authors have provided the objectives and requirements of information collection with a taxonomy of various data gathering techniques. In addition, the authors have reviewed various traditional collection tools, nodes, and mechanisms in terms of security-related and data collection based on proposed objectives and requirements towards high-quality related security. Further, the authors have proposed various open challenges by concluding the paper with suggested future directions. Wu et al. [22] have proposed an analysis-based architecture for big data secure clustering management for the control planes. The authors have proposed an authentication mechanism for managing the clusters and enabling the data analysis using a colony optimization scheme. The comparative and simulated results increased the feasibility and efficiency of the proposed framework in control planes. Zhou et al. [23] have proposed a differently private scheme to preserve the data among edge nodes and users while communicating the information in various networks. In order to guarantee the trusted computing, a trust-based mechanism is proposed to evaluate the end user's reliability. The experimental results supported the increasing multimedia big data information while striking the balance among trustworthy, privacy-preserving and caching multimedia contents and big data collection prediction. Han et al. [24] have investigated an agile confidential transmission strategy for securing the big data information transmission. The authors have combined the opportunistic beamforming and driven cluster schemes for managing the huge data collection from various base stations. For the purpose of a secure and confidential transmission among clusters, the authors have combined the beamforming and driven clusters for reliable and efficient changing in network environment. In order to evaluate and validate the proposed scenario, the results have performed average secrecy of sum capacity and number of authorized users while accessing the systems.

Though numbers of approaches have been proposed by various researchers and scientists alone, very few of them have focused on an IoT-based artificial network to process and compute big data by ensuring a secure communication multihoming network. The number of smart techniques that can be helpful for monitoring the data processing, transmission, and communication of big data in multihoming by monitoring their nodes is still unexplored in the literature. In addition, the integration of automated schemes for

managing and processing big data is considered one of the significant research interests. Further, the integration of intelligent techniques while analysing the network algorithms, security, and clustering of multiple networks having various protocols, configurations, and attributes may benefit in a number of techniques. The automated systems may further need to manage storing the huge amount of information from heterogenous networks in an efficient way. The aim of this paper is to propose an efficient IoT-based artificial network to process and compute big data by ensuring a secure communication multihoming network approach with optimum evaluation results using the Bayesian Rule (BR) and Levenberg-Marquardt (LM) algorithms [25]. In addition, the proposed scheme is validated through a set of synthesized datasets against various monitoring and value processing results.

2.1. Proposed Approach. Presently, a number of machine learning, trust-based, and artificial intelligence algorithms have been proposed by various researchers and scientists. The smart and AI-based schemes in IoT benefitted in various phases of application while ensuring a secure transmission or communication among nodes in the network. In this paper, we have proposed an IoT-based artificial network to process and compute big data by ensuring a secure communication multihoming network. An IoT-based ANN is termed as an automated computational and processing scheme inspired by numerous neurons based on the concept of the biological neural network. The general definition of a neuron is defined as the lexeme cell as an assortment of numerous biological neurons referred to as the base for modelling an automated AI-based architecture [26]. An IoT-based ANN is defined as a mathematical model for processing the classification of data, nonlinear function, and regression schemes. It is capable of generating an automated decision model via multilayered perceptron. Figure 1 presents a multistage perceptron IoT-based ANN architecture having input through various smart sensors, hidden layers used for processing and computing the inputs, and an output layer used to generate the final output depending upon the provided input. An IoT-based ANN consists of a set of "o" number of outputs, H_h number of hidden/middle layers, and I_i number of inputs as defined in

$$\alpha_r(t) = \sum_{\alpha=1}^{H_h} W^2{}_{rs} F(.) \sum_{\gamma_{=1}} I_i W^1{}_{\alpha r} \alpha_s(t)^0 + b^1{}_{\alpha}, \quad \text{where } 1 \le r \le 0, \tag{1}$$

where W_{rs} and W_{ar} denote the connection of edges via weights among input, middle, and output layers. In addition, the function $F(.)$ in this equation represents an activation function (AF) that is defined as a sigmoid function to determine an appropriate processing and computation of trust values by evaluating the probabilities using the ANN algorithm. Further, the values in W_{rs} and W_{ar} denote an appropriate scheme using the Levenberg-Marquardt (LM) and Bayesian Rule (BR) principle for an optimum and efficient mechanism. The involvement of the AI-based scheme in the IoT system for processing or securing the multihoming network data can further benefit the system in a variety of ways. The following sections detailed the LM and BR classification for generating an accurate analysis of output results.

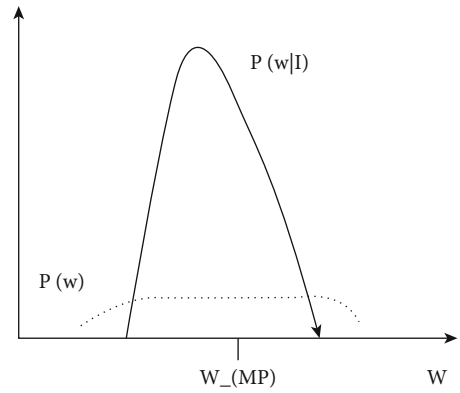


Figure 2: Prior to posterior weight changing process.

2.1.1. Levenberg-Marquardt (LM) Algorithm. LM is a deterministic and gradient-based local optimum algorithm. The benefit of using the LM algorithm while training the multistage perceptron architecture is the fast and average convergence rate by providing the stability in the system. Similar to the quasi-Newton scheme, LM was developed for a second-order derivation training speech approach without computing the Hessian matrix. The Hessian matrix is approximated while performing the function of sum of squares as

$$H_M = Q^T Q, \tag{2}$$

where the gradient can be evaluated as

$$G = Q^T \sigma, \tag{3}$$

where Q is defined as the Jacobian matrix containing the first derivations of error with respect to biases and weights. In addition, σ denotes the vector of errors in a network. The Jacobian matrix can be evaluated using a standard BR technique where the expected outputs through hidden layers are represented as

$$\alpha_q(t) = F'(I_i(t)) \sum_q \sigma^r{}_q(t) W^2{}_{rq}(t-1), \tag{4}$$

where q is the number of hidden layer neurons having r number of layers. Further, the LM algorithm uses the approximation to the Hessian matrix as

$$\text{delta}w = -\left[Q^T Q + \mu I\right]^{-1} Q^T \sigma, \tag{5}$$

where w represents the differential weights and μ denotes the controlling parameters. Whenever the mu is scaled to zero, then it is defined as Newton's method using the Hessian matrix approximation. However, when mu is large, then it becomes gradient descent having small step size. Newton's

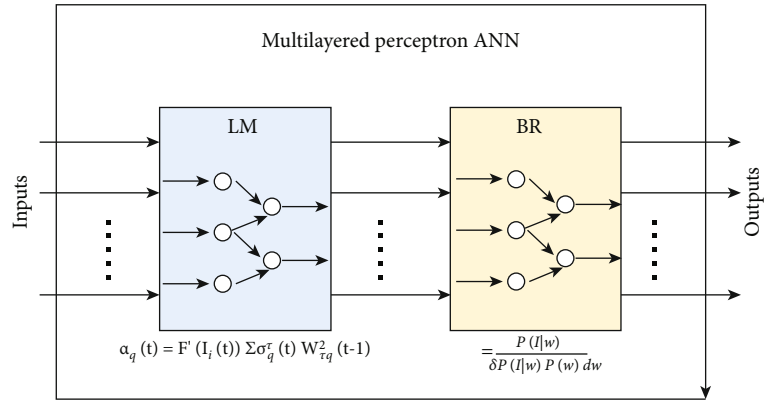


Figure 3: Prior to posterior weight changing process.

method is much accurate and faster near an error minimum. Therefore, the value of mu is decreased after every successful process and increased only when the step increases the performance function.

2.1.2. Bayesian Rule (BR) Algorithm. Later on, the LM algorithm is integrated with the BR method in order to further optimize the processed data. The BR algorithm is defined as

$$P(x \mid I) = \frac{P(I \mid x)}{P(I)}, \tag{6}$$

where $P(x)$ represents the prior probability of parameter x before actually seeing the processed information and $P(x \mid I)$ represents the likelihood where the probability of information I is. The BR was basically used to illustrate the posterior probability of x given the information I. In common, BR provides an entire distribution over all possible x values. This process was applied to a neural network by coming up with the probability distribution over weights w upon giving the training data as $P(w \mid I)$.

The posterior distributions upon weights are determined as

$$P(w \mid I) = \frac{P(I \mid w)P(w)}{P(I)}, \tag{7}$$

$$P(w \mid I) = \frac{P(I \mid w)}{\delta P(I \mid w)P(w)dw}. \tag{8}$$

Further, in the BR rule formulation, the learning of weights changes the beliefs about prior $P(w)$ and posterior $P(w \mid I)$ weights as consequences of seeing the information represented in Figure 2. As depicted in Figure 2, the weights of the learning rates are changes as per the information received and processed from various inputs. The inputs received from malicious nodes are analysed through their energy consumption and distribution ratio in the network. The nodes having malicious behaviour will always process false or alternate information with a number of generated errors.

Table 2: Simulation parameters.

Multilayered network	Training (%)	Testing (%)	Time (secs)
Bayesian Rule (BR)	63.52	36.48	8.562
Levenberg-Marquardt (LM)	66.89	33.11	2.598
IoT nodes	150	50	60 sec

Table 3: Synthesized simulated results.

Class	Proposed mechanism	Basic mechanism
Specificity	0.081	0.075
Sensitivity	0.912	0.0873
Accuracy	98.58	97.46
F-measure	1.19	1.10
ROC (receiver operator characteristic)	0.87	0.83

3. Working of the Proposed Approach Using LM and BR Algorithms

The working of the proposed mechanism using the BR and LM algorithm is explained through a diagram as presented in Figure 3. The above-mentioned BR and LM algorithms are used for ensuring a secure and efficient communication transmission while processing the data. The input in terms of received signals/information is passed into the ANN multilayered perceptron. Initially, the LM algorithm is applied on the inputs for computing the convergence rate and weights (trust) of each node's input while mentioning the error. Each node including hidden nodes will evaluate the gradient and Jacobian matrix. The errors while analysing the weights from various input nodes will be handled using the controlling parameters as depicted in equations (4) and (5). In addition, Newton's method is further analysed to ensure the fast and accurate results while minimizing the errors.

Now, in order to ensure an efficient processing and computation of weights after analysing or computing the weights from each node, the BR algorithm is applied over LM to

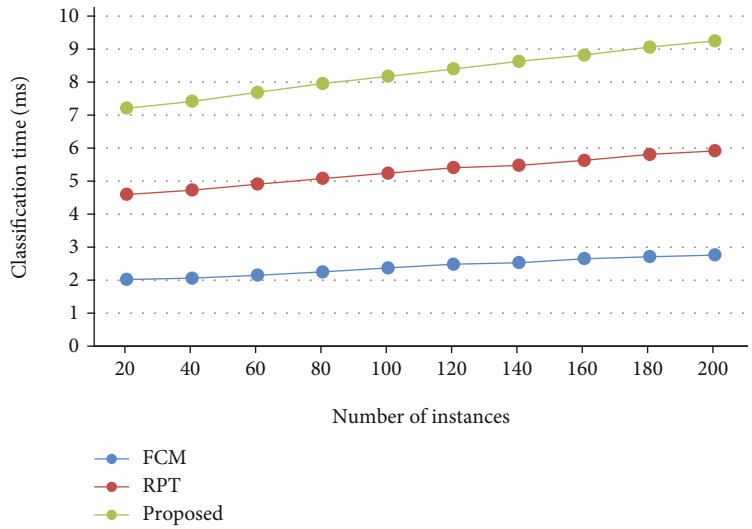


Figure 4: Classification time.

optimize the processed or recorded information from the inputs. The input is probabilistically distributed to the various nodes for computing and processing the efficient distribution of information while stabilizing the system. The posterior distribution of weights using the BR algorithm is computed using equations (7) and (8).

3.1. Performance Analysis. The proposed IoT-based artificial network, to process and compute big data by ensuring a secure communication multihoming network mechanism, is validated and experimented over an existing smart mechanism against several security threats. The proposed phenomenon is analysed over a synthesized dataset with a conventional smart-based scheme where the number of instances or inputs to the network is taken as 20-200 using a MATLAB simulator. The number of instances is input in the network where BR and LM algorithms are used to analyse and process the incoming data. Table 2 depicts the measured simulation results with several analysed values where the proposed approach having BR and LM algorithms is used to process the information. The number of input information is passed through both the mechanisms where the data is divided into training and testing analysis for optimizing or processing the information. Further, the input is probabilistically distributed to the various nodes for computing and processing the information distribution while stabilizing the system. The weights of posterior distribution using the BR algorithm is further computed using various equations.

The simulated results are analysed over various security measures as follows:

Accuracy: the accuracy is defined as the number of values required to produce accurate results.

Specificity: it is defined as the false-positive rates to categorize the weights of each node that are designed incorrectly. In the multihoming network where the data is recorded and processed from various networks, the probability of false-positive rates may be very high.

F-measure and ROC (receiver operator characteristic): they are used to determine and illustrate the classification accuracy of the proposed phenomenon. It is used to compute the *F*1 score of each node by recognizing precision and recall. The classification accuracy measures the efficient and trusted behavior of the network while processing various inputs of heterogenous networks.

Sensitivity: it determines the true-positive results which are correctly recognized by the system.

The synthesized simulation results are represented in Table 3.

3.2. Baseline Approach. For analysing the performance measure, the proposed phenomenon is compared against a baseline approach proposed by Wu et al. [22] which generated an analysis-based architecture for big data secure clustering management for the control planes. The authors have proposed an authentication mechanism for managing the clusters and enabling the data analysis using a colony optimization scheme. The comparative and simulated results increased the feasibility and efficiency of the proposed framework in control planes. In addition, the comparative results are analysed over various AI-based schemes such as Fuzzy *C* Means (FCM) and REPTree (RPT) methods. The proposed mechanism is analysed against various existing decision-making schemes to measure the accuracy and security of the processed information in multihoming networks.

4. Results and Discussion

Numbers of classified algorithms are evaluated and analysed based upon two statistical values, namely, accuracy and time of classified values. The classification time as shown in Figure 4 of the AI-based scheme is analysed over the existing mechanism against various data formats. Figure 4 shows the classification time that shows better results as compared to the existing scheme. The classification time of the proposed approach is better as compared to the existing AI-based schemes because of the optimized and discrete LM algorithm that trains the multistage perceptron architecture in a faster and average convergence rate by providing the stability in

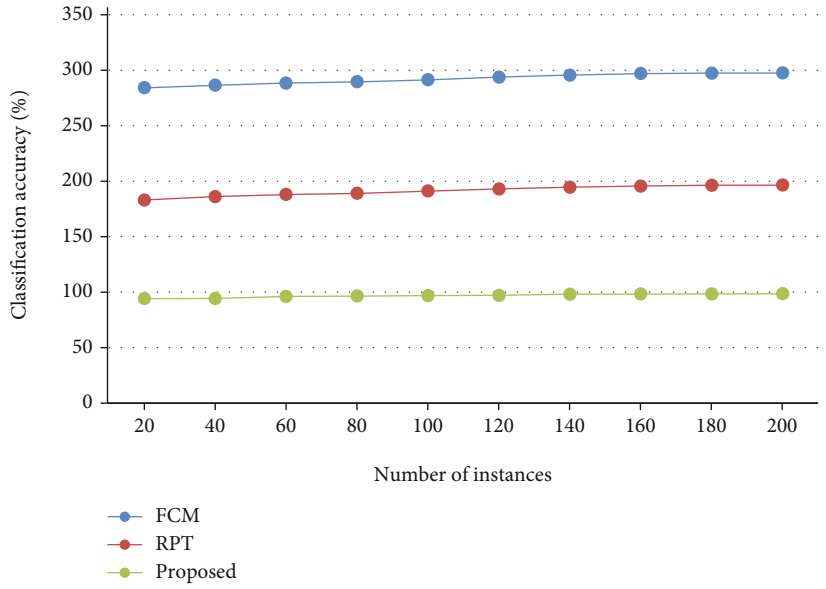


Figure 5: Classification accuracy.

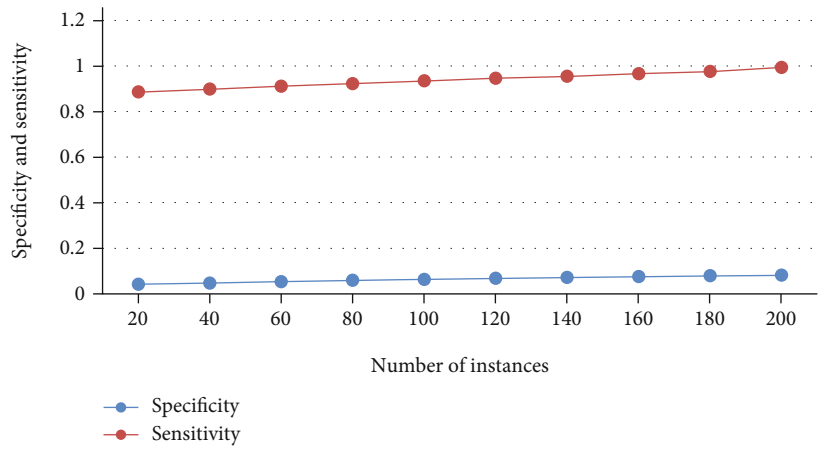


Figure 6: Specificity and sensitivity.

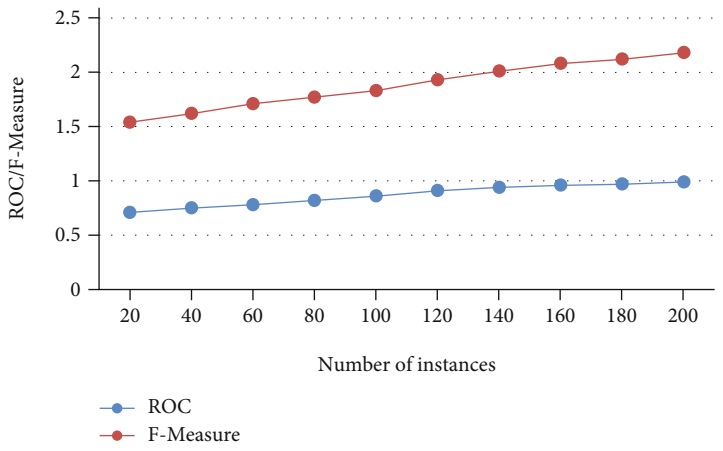


Figure 7: *F*-measure and ROC.

the system. In addition, Figure 5 depicts the classification accuracy of the proposed scheme that performs in a relative manner to the existing mechanism.

The accuracy of the proposed phenomenon is approximately 98% which is improved as compared to that of the existing scheme. The significant outperformance of the proposed phenomenon is due to the BR method that optimizes the processed data to improve the accuracy of the overall system in the network. Figure 6 outperforms while monitoring and analysing the activities of each and every individual. Further, Figure 6 represents the specificity and sensitivity of the proposed phenomenon using AI-based architecture where the values are analysed and processed over smart devices. As depicted in Figure 6, the specificity and sensitivity of the proposed record are improved as a huge amount of information from various networks is processed via the LM mechanism that may enhance the multistage architectural processing, management, and security of records in multihoming networks. Furthermore, Figure 7 represents the ROC and *F*-measure of the proposed phenomenon to determine the accuracy and monitoring against various existing schemes. Figure 7 determines the optimum value results of the proposed approach. The significant improvement of the proposed phenomenon as compared to the existing scheme is due to BR and LM schemes that manage and optimize the huge number of generated records from various types of networks into a single environment.

The processing of big data is optimized through a deterministic and gradient-based local optimum algorithm while training the multistage perceptron architecture through a fast and average convergence rate by providing the stability in the system.

5. Conclusions

This paper proposes an AI-based secure multihoming mechanism for ensuring a secure transmission and processing of big data using the Bayesian Rule (BR) and Levenberg-Marquardt (LM) algorithms. For an efficient monitoring and processing of big data risks while communicating, the LM and BR mechanisms processed the various inputs from heterogenous networks and analyse the weights of each node. The hybrid of the LM and BR algorithm in multihoming networks ensured the efficiency and security while processing the huge information from various networks. The proposed approach efficiently processed the classification of data, nonlinear function, accuracy, and regression schemes in multihoming networks. In addition, the proposed mechanisms are capable of generating an automated decision model via multilayered perceptron using the hybrid of LM and BR schemes. Further, the proposed phenomenon significantly processes and monitors the processed data while proving the security with optimal time delay. The validity and verification of the proposed scheme are experimented over various simulating results against various monitoring and processing parameters such as accuracy, specificity, sensitivity, *F*-measure, and ROC.

The number of automated controlling schemes such as explainable artificial intelligence to further analyse the huge processed records in an efficient and secured manner in multihoming networks by monitoring their activities can be considered in future communication. In addition, the concept of backpropagation in the proposed mechanism is not considered at this stage where instead of computing the gradient of lost function, we have calculated the error propagation ratio and accuracy from the input information in the network.

Data Availability

This paper does not need any online data for the simulation. The present study is based on synthesized data generated randomly by the authors based on some parameters mentioned in the above text.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

This work is supported by the Netaji Subhas University of Technology, India, and University of the Fraser Valley, Canada. The work was funded by the Abu Dhabi University (Faculty Research Incentive Grant 19300483—Adel Khelifi), United Arab Emirates (https://www.adu.ac.ae/research/research-at-adu/overview).

References

[1] S. Madden, "From databases to big data," *IEEE Internet Computing*, vol. 16, no. 3, pp. 4–6, 2012.

[2] Z. S. Ageed, S. R. Zeebaree, M. M. Sadeeq et al., "Comprehensive survey of big data mining approaches in cloud systems," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 29–38, 2021.

[3] K. A. Bryan and J. S. Gans, "A theory of multihoming in rideshare competition," *Journal of Economics and Management Strategy*, vol. 28, no. 1, pp. 89–96, 2019.

[4] C. Cennamo, H. Ozalp, and T. Kretschmer, "Platform architecture and quality trade-offs of multihoming complements," *Information Systems Research*, vol. 29, no. 2, pp. 461–478, 2018.

[5] A. Sharma, G. Rathee, R. Kumar et al., "A secure, energy- and SLA-efficient (SESE) E-healthcare framework for quickest data transmission using cyber-physical system," *Sensors*, vol. 19, no. 9, pp. 2011–2119, 2019.

[6] Q. Xu, Z. Su, Q. Zheng, M. Luo, B. Dong, and K. Zhang, "Game theoretical secure caching scheme in multihoming edge computing-enabled heterogeneous networks," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4536–4546, 2018.

[7] Z. Zhang, "Artificial neural network," in *Multivariate Time Series Analysis in Climate and Environmental Research*, pp. 1–35, Springer, Cham, 2018.

[8] G. Rathee, S. Garg, G. Kaddoum, and B. J. Choi, "A decision-making model for securing IoT devices in smart industries," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4270–4278, 2020.

[9] J. Arshad, M. A. Azad, R. Amad, K. Salah, M. Alazab, and R. Iqbal, "A review of performance, energy and privacy of

intrusion detection systems for IoT," *Electronics*, vol. 9, no. 4, pp. 1–24, 2020.

[10] G. Rathee, H. Saini, and G. Singh, "Aspects of trusted routing communication in smart networks," *Wireless Personal Communications*, vol. 98, no. 2, pp. 2367–2387, 2018.

[11] H. Gao, Y. Xu, X. Liu et al., "Edge4Sys: a device-edge collaborative framework for MEC based smart systems," in *2020 35th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pp. 1252–1254, New York, NY, USA, 2020.

[12] G. Rathee, M. Balasaraswathi, K. P. Chandran, S. D. Gupta, and C. S. Boopathi, "A secure IoT sensors communication in industry 4.0 using blockchain technology," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 533–545, 2021.

[13] J. S. Warm, W. N. Dember, and P. A. Hancock, "Vigilance and workload in automated systems," in *Automation and Human Performance: Theory and Applications*, pp. 183–200, CRC Press, 2018.

[14] D. Ferraioli, A. Meier, P. Penna, and C. Ventre, "Automated optimal OSP mechanisms for set systems," in *International Conference on Web and Internet Economics*, pp. 171–185, Springer, Cham, 2019.

[15] R. Mamlook, O. F. Khan, M. M. Haddad, H. S. Koofan, and S. M. Tabook, "Controlling future intelligent smart homes using wireless integrated network systems," *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 15, no. 2, 2017.

[16] E. G. de Santerre, S. Jammoul, and L. Toutain, "Solving the ingress filtering issue in an IPv6 multihomed home network," in *Ninth International Conference on Networks*, pp. 272–278, Menuires, France, 2010.

[17] L. Wang, H. Guo, Y. Su, and C. Liu, "A reactive learning mechanism for multihoming MN on PMIPv6," in *2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010)*, vol. 3, pp. 76–79, Wuhan, China, 2010.

[18] J. Bi, P. Hu, and L. Xie, "Site multihoming: practices, mechanisms and perspective," in *Future Generation Communication and Networking (FGCN 2007)*, vol. 1, pp. 535–540, Jeju, Korea (South), 2007.

[19] L. He, Z. Yan, and M. Atiquzzaman, "LTE/LTE-a network security data collection and analysis for security measurement: a survey," *IEEE Access*, vol. 6, pp. 4220–4242, 2018.

[20] F. Li, R. Xie, Z. Wang et al., "Online distributed IoT security monitoring with multidimensional streaming big data," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4387–4394, 2020.

[21] H. Lin, Z. Yan, Y. Chen, and L. Zhang, "A survey on network security-related data collection technologies," *IEEE Access*, vol. 6, pp. 18345–18365, 2018.

[22] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, "Big data analysis-based secure cluster management for optimized control plane in software-defined networks," *IEEE Transactions on Network and Service Management*, vol. 15, no. 1, pp. 27–38, 2018.

[23] P. Zhou, K. Wang, J. Xu, and D. Wu, "Differentially-private and trustworthy online social multimedia big data retrieval in edge computing," *IEEE Transactions on Multimedia*, vol. 21, no. 3, pp. 539–554, 2018.

[24] S. Han, S. Xu, W. Meng, and C. Li, "An agile confidential transmission strategy combining big data driven cluster and OBF," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10259–10270, 2017.

[25] J. Adnan, N. N. Daud, S. Ahmad et al., "Heart abnormality activity detection using multilayer perceptron (MLP) network," in *AIP Conference Proceedings (Vol. 2016, No. 1, p. 020013)*, AIP Publishing LLC, 2018.

[26] F. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble model of artificial neural networks with randomized number of hidden neurons," in *IEEE 8th International Conference on Information Technology in Asia (CITA)*, pp. 1–5, Kota Samarahan, Malaysia, 2013.

[27] G. Rathee, S. Garg, G. Kaddoum, Y. Wu, D. N. K. Jayakody, and A. Alamri, "ANN assisted-IoT enabled COVID-19 patient monitoring," *IEEE Access*, vol. 9, pp. 42483–42492, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9977751, 15 pages
https://doi.org/10.1155/2021/9977751

Research Article

An Intelligent Big Data Management System Using Haar Algorithm-Based Nao Agent Multisensory Communication

Fatmah Abdulrahman Baothman

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21431, Saudi Arabia

Correspondence should be addressed to Fatmah Abdulrahman Baothman; fbaothman@kau.edu.sa

Received 4 March 2021; Accepted 4 May 2021; Published 14 July 2021

Academic Editor: Nawab Muhammad Faseeh Qureshi

Artificial intelligence (AI) is progressively changing techniques of teaching and learning. In the past, the objective was to provide an intelligent tutoring system without intervention from a human teacher to enhance skills, control, knowledge construction, and intellectual engagement. This paper proposes a definition of AI focusing on enhancing the humanoid agent Nao's learning capabilities and interactions. The aim is to increase Nao intelligence using big data by activating multisensory perceptions such as visual and auditory stimuli modules and speech-related stimuli, as well as being in various movements. The method is to develop a toolkit by enabling Arabic speech recognition and implementing the Haar algorithm for robust image recognition to improve the capabilities of Nao during interactions with a child in a mixed reality system using big data. The experiment design and testing processes were conducted by implementing an AI principle design, namely, the three-constituent principle. Four experiments were conducted to boost Nao's intelligence level using 100 children, different environments (class, lab, home, and mixed reality Leap Motion Controller (LMC). An objective function and an operational time cost function are developed to improve Nao's learning experience in different environments accomplishing the best results in 4.2 seconds for each number recognition. The experiments' results showed an increase in Nao's intelligence from 3 to 7 years old compared with a child's intelligence in learning simple mathematics with the best communication using a kappa ratio value of 90.8%, having a corpus that exceeded 390,000 segments, and scoring 93% of success rate when activating both auditory and vision modules for the agent Nao. The developed toolkit uses Arabic speech recognition and the Haar algorithm in a mixed reality system using big data enabling Nao to achieve a 94% success learning rate at a distance of 0.09 m; when using LMC in mixed reality, the hand sign gestures recorded the highest accuracy of 98.50% using Haar algorithm. The work shows that the current work enabled Nao to gradually achieve a higher learning success rate as the environment changes and multisensory perception increases. This paper also proposes a cutting-edge research work direction for fostering child-robots education in real time.

1. Introduction

Artificial intelligence (AI) was introduced half a century ago. Researchers initially wanted to build an electronic brain equipped with a natural form of intelligence. The concept of AI was heralded by Alan Turing in the 1950s, who proposed the Turing test to measure a form of natural language (symbolic) communication between humans and machines. In the 1960s, Lutfi Zadah proposed fuzzy logic with dominant knowledge representation and mobile robots [1]. Stanford University created the Automated Mathematician to explore new mathematical theories based on a heuristic algorithm. However, AI had become unpopular in the 1970s due to its inability to meet unrealistic expectations. The 1980s offered a promise for AI as sales of AI-based hardware and software for decision support applications exceeded $400 million [2]. By the 1990s, AI had entered a new era by integrating intelligent agent (IA) applications into different fields, such as games (Deep Blue, which is a chess program developed at Carnegie Mellon that defeated the world champion Garry Kasparov in 1997), spacecraft control, security (credit card fraud detection, face recognition), and transportation (automated scheduling systems) [3–7]. The beginning of the 21st century witnessed significant advances in AI in

industrial business and government services with several initiatives, such as intelligent cities, intelligent economy, intelligent industry, and intelligent robots [3].

A unified definition of AI has not yet been offered; however, the concept of AI can be built from different definitions:

(i) It is an interdisciplinary science because it interacts with cognitive science

(ii) It uses creative techniques in modeling and mapping to improve average performance when solving complex problems

(iii) It implements different processes to imitate intelligent human or animal behavior. Fourth, the developed system is either a virtual or a physical system with intelligent characteristics

(iv) It attempts to duplicate human mental and sensory systems to model aspects of "humans" thoughts and behaviors

(v) It passes the intelligence test if it interacts completely with other systems or creatures worldwide and in real time

(vi) It follows a defined cycle of sense–plan–act

The present study proposes the definition of AI as follows: "AI is an interdisciplinary science suitable for implementation in any domain that uses heuristic techniques, modeling, and AI-based design principles to solve complex problems. Single or combined processes in perceiving, reasoning, learning, understanding, and collaborating can improve system behavior and decision-making. The goal of AI is to enable virtual and physical intelligent agents, including humans and/or systems that continuously upgrade their intelligence to attain superintelligence. Agents should be able to integrate with one another in fully learning, teaching, adapting themselves to dynamic environments, communicating logically, and functioning efficiently with one another or with other creatures in the world and real time through sense–plan–act–react cycles." The three-constituent principle for an agent suggests that "designing an intelligent agent involves constituents, the definition of the ecological niche, the definition of the desired behaviors and tasks, and design of the agent [8, 9]." Therefore, an agent's intelligence can grow in time using the "here and now" perspective during interactions in different dynamic environments. In the present study, the robot agent Nao's design is not among the required tasks, but the other two constituents are related to the environment and involve interactions with a human agent. Therefore, this work defines the ecological niche using different environments (a classroom, a lab, and a home), focusing on a mixed-reality environment. Nao's functions are present according to the desired behavior as teaching simple mathematics to a child. The objective is to improve Nao's learning ability and increase its intelligence. Thus, the study shows that "the three-constituents principle, the definition of the ecological niche, and the definition of the desired behaviors and tasks [2]" are sufficient to increase Nao's intelligence.

(i) The "here and now" perspective: related to three-time frames and shows that the behavior of any 'agent's system matures over a certain period and is associated with three states

(ii) State-oriented: describes the actual mechanism of the agent at any instance of time

(iii) Learning and development: relates to learning and development from state-oriented action

(iv) Evolutionary: explains the emergence of a higher level of cognition through a phylogenetic perspective by emphasizing the power of artificial evolution and performing more complex tasks

The Mixed Reality System TouchMe provides a third-person camera view of the system instead of human eyes [10]. The third-person camera view is considered more efficient for inexperienced users to interact with the robot [11]. Leutert et al. [12] reported using augmented spatial reality, a form of mixed reality, to relay information from the robot to the user's workspace. They used a fixed-mobile projector. Socially aware interactive playgrounds [13] use various actuators to provide feedback to children. These actuators include projectors, speakers, and lights. These "interactive playgrounds can be placed at different locations, such as schools, streets, and gyms. Humans produce, interpret, and detect social signals (a communicative or informative signal conveyed directly or indirectly) [13]." Thus, their social signals can be used to enhance interactions with others. Various studies have been conducted on teaching humans to use robots in various environmental settings. RoboStage module implements learning among junior high school students through mixed reality systems [14–20]. Its creators compared the use of physical and virtual characters in a learning environment. RoboStage enables module interactions in robots to use voice and physical objects to achieve three stages of events: learning, situatedness, and blended. These events help students learn and practice activities, understand an environment, and execute an event. GENTORO uses a robot and a handheld projector to interact with children and perform a storytelling activity [21–27]. Its creators studied the effect of using a small handheld projector on the storytelling process. They also discussed the effects of using audio interactions instead of text and a wide-angle lens.

The agent matures into an adult by which the process in any state is affected by its previous state. The present study has focused on state-oriented and learning and development states to observe its outcome in association with the evolutionary state [28–30]. The proposed definition enhances research at the experimental design level using multisensory technologies to improve intelligence interaction and growth by applying the AI design principle [31–33]. Enhanced interaction between humans and robots improves learning, especially in the case of a child. Motion and speech sensor nodes are fused to this end. Contemporary children are familiar with handheld devices such as mobile phones, tablets, pads, and virtual reality cameras. Therefore, the toolkit developed in this study uses a mixed reality system featuring different

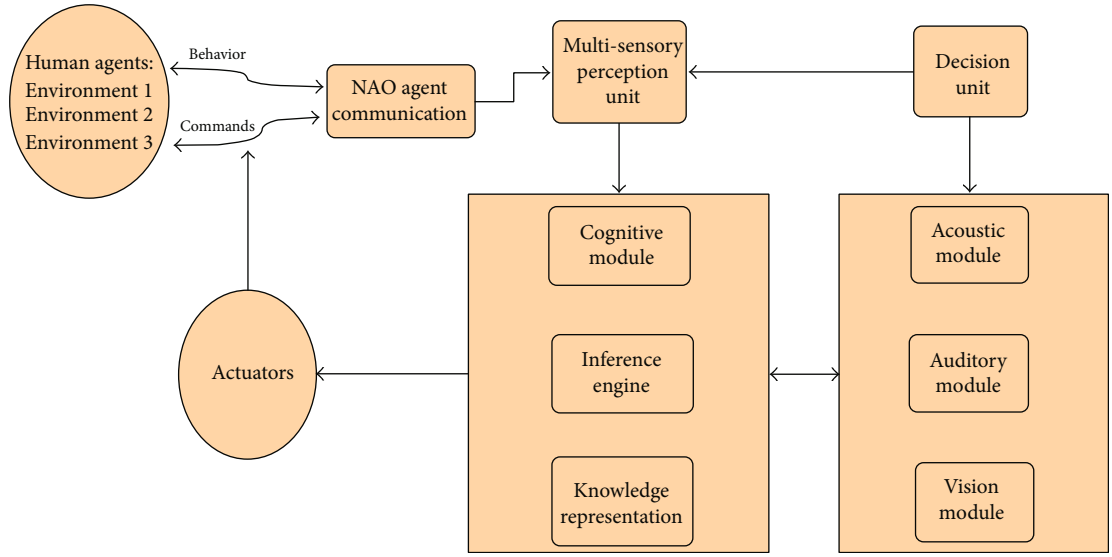


Figure 1: Nao agent multisensory perception communication concept design. Preliminary training datasets for Nao involved https://github.com/IBM/watson-nao-robot, Andreas Ess' webpage (ethz.ch), The NAL dataset (http://inria.fr), SpeechRecognition · PyPI, speech corpus [34], and Leap Motion Controller [35].

ways of interaction between a child and a robot agent. This study makes the following contributions.

(i) Enhancing the humanoid robot Nao's learning capabilities with the objective to increase the robot's intelligence, using a multisensory perception of vision, hearing, speech, and gestures for HRI interactions

(ii) Implementing Arabic Speech Agent for Nao using phonological knowledge and HMM to eventually activate child-robot communication [34]

(iii) It developed a toolkit using Arabic speech recognition and the Haar algorithm for robust image recognition in a mixed reality system architecture using big data enabling Nao to achieve a 94% success learning rate featuring different environments, and for LMC, the highest accuracy of 98.50% using the Haar algorithm

The remainder of the study is organized mainly into Materials, Data, and Methods, which describe the architecture and experiment design, while the Discussion and Results section covers the intelligent big data management system using Haar algorithm-based Nao Agent Multisensory Communication in mixed reality and using LMC. Finally, the Conclusion and Future Work of the proposed study.

2. Materials, Data, and Methods

The experiment initiated at King Abdul-Aziz University with an Aldebaran representative was related to a three-year-old robot Nao, which could not speak Arabic or solve simple mathematics. The study analysis was initiated by selecting the suitable artificial intelligence principle design for the study. The experiment's goals and tasks were defined precisely to increase Nao's intelligence to at least seven years old. The Nao mathematics intelligence measurements were based on solving 100 children's exercises for basic addition, subtraction, and multiplication problems with human agents' help. Nao also reached the level of understanding simple sentences for Arabic language speech recognition. The experiment time scale was set for a total of two years. The study is aimed at involving the robot Nao in the learning-teaching process using interaction and multisensory Nao agent perceptions by exposing Nao to different environments (see Figure 1), enabling communication concept design. However, the present work focused more on the mixed reality environment.

The data collection involved two agents, the Nao robot and the user. For the user speech datasets (1) is collected, the recordings were for spoken Arabic numbers from 0 to 9. A total of five male and female speakers were asked to pronounce each number three times. All speakers were from Jeddah city, and the recordings were conducted in an ordinary quiet room. The speech data were captured at 16 KHz at an average speaking rate. For each person, the recording session lasted 60-90 minutes. The acoustic data were transmitted into a war sound recording file for later analysis. The Nao voice recognition consists of recording (ALProxy module ALAudioRecorder) and recognition modules tested by asking each person to repeat the number until the recognizer gets it correct. Python programming language is used. The environment Windows, Python IDLE (Python GUI), and NAOqi operating system using Choregraphe modules such as Almath and Python's Automatic Speech Recognition library were implemented. The computer is a static environment for data and processing and computational analysis.

For image collection, the hand dataset (2) is generated at KAU class, lab, home, and mixed reality; the author used The NAL dataset (http://inria.fr) for initial training and gathered the dataset by implementing the Leap Motion Controller

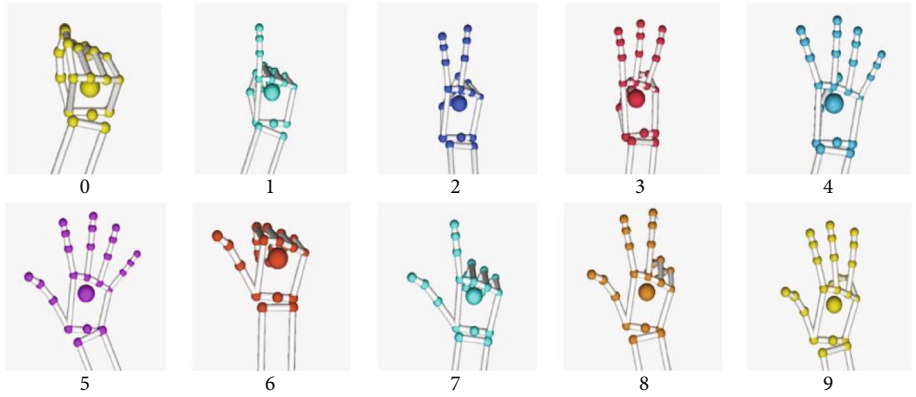


Figure 2: Visualization of gestures tracked by LMV [35].

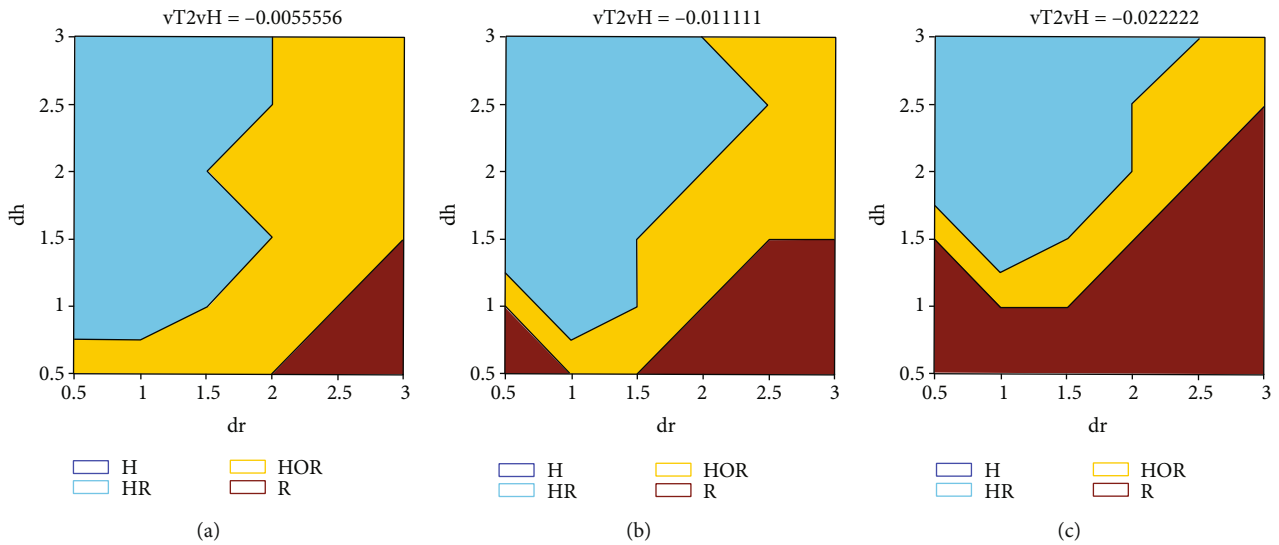


Figure 3: Time cost increases from graph (a) to graph (c) [36].

(LMC) Visualizer software for hand tracking for ten number gestures acquired from 34 subjects of three age groups (six and thirty) [35]. Figure 2 shows the gesture numbers in Leap Motion Visualizer (LMV) during hand tracking while LMV extracts the finger features based on path (x, y, z) and place (x, y, z) for thumb, index, middle, ring, and pinky.

NaoQi supports C++, Python, and JavaScript programming languages to be used on the robot. Several built-in modules include auditory, vision, and recognition. For example, the ALModule API has three methods to changeDatabase(), getParam(), and setParam() in C++. The Haar algorithm module was written in Python and applied to improve and stimulate Nao's vision. A snapshot model of a child showing fingers indicated a number. The Nao agent stores the number represented by a human agent in a database to improve its learning capabilities. Nao senses environment via sonar or pumper sensors and cameras; together, they support the Nao agent perceptions in addition to a trajectory algorithm. According to [36], "The human's sensitivities $(\dot{d}\ h)$ and the robot $(\dot{d}\ r)$ are ranged along X and Y axes human dominance with robot interaction reduces as the time cost increases from a graph 'a' $(vT2vH = -0.0055)$ to graph 'c' $(vT2vH = -0.0222)$, and consequently, the collaboration reduced." The following (Figure 2) represents the best performance level and shows that the decrease "from 92% in graph 'a' to 60% in graph 'c'" [36] in human dominance collaboration level as time cost and human response time increase, as indicated by the HR and HOR areas (see Figure 3).

In this work, an objective function is implemented as a collaborative model describing system performance within a specific environment; [19] used only four parameters (hits, false alarms, and missing target items) for a given process. To fit the experiments' objective function, the author added two more parameters to improve Nao's learning experience and robot-human interaction in different environments. There are two agents' interactions, a human and a robot, each could score a process, with a defined task, in four specified environments. Therefore, the author defined the objective functions (1) for both agents by six parameters, rather than four, to measure the system interaction performance ${}^{O}SP$ as follows:

$$\begin{aligned} {}^{O}SP = {}^{O}HAgInter + {}^{O}RobInter + {}^{O} \\ HAgtask + {}^{O}RobAgtasks + {}^{O}RobEnv + {}^{O}Ts, \end{aligned} \tag{1}$$

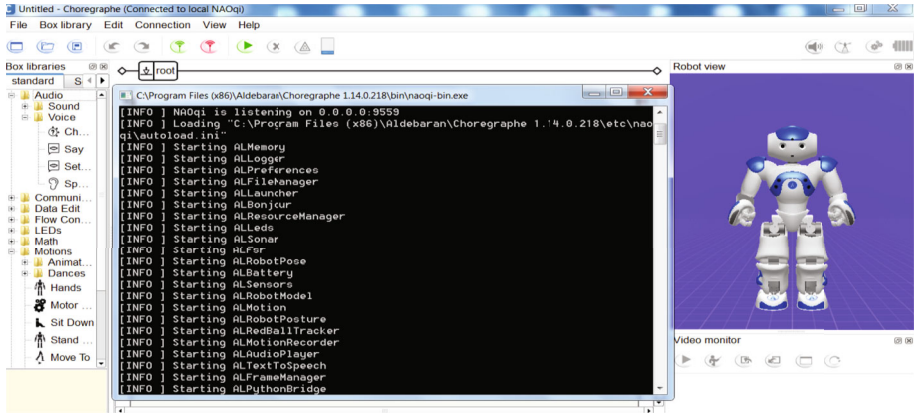


Figure 4: Nao agent initiation

where *OHAgInter* is the human information interaction by (0,1), *ORobAgInter* is the robot interaction by (0,1), ${}^{O}HAgt\ ask$ is the number of tasks by a human (1,2,*n*), ${}^{O}RobAgtask\ s$ is the number of tasks by a Nao robot (1,2,*n*), ${}^{O}RobEnv$ is the robot environment (1 = class, 2 = lab, 3 = home, 4 = mixed reality), and ${}^{O}Ts$ represents the time operational cost.

A loss function is a part of a cost function which is a type of objective function. In this work, the author calculated both objective and cost functions only. In function (2), the system operational time cost (${}^{O}{}_{Ts}$) measures the activities' cost while detecting a true hit or a false miss for a target item. For example, the operation time for image processing by Nao to execute true hit in recognizing a number or counts false miss if wrong. The operational costs here are counted for two agents while they interact in different environments. It involves the cost of time spent within the operation, the decision time by the two agents (human/Nao robot) for identifying whether an item is a hit or not. The default cost value is chosen to be 4 seconds since two agents (Nao robot/human) are operating simultaneously. This work is calculated as follows:

$$O_{Ts} = t_{HS} \cdot C_t + [N \cdot P_{HMS} \cdot p_{Hs} + N \cdot (1 - P_S) \cdot P_{FHs}] \cdot O_c + t_{RS} \cdot C_t + [N \cdot P_{rMS} \cdot P_{rs} + N \cdot (1 - P_S) \cdot P_{Frs}] \cdot O_c. \quad (2)$$

Following [36], *tHs* and t_{RS} are the time needed to execute a task by a human and a Nao robot; ${}^{O}{}_{C}$ is the required operation cost for recognizing a solo item during the interactive mode, *Ct* represents the cost of a time unit for identifying an item, *N* represents the number of items for image detection, and P_{HS} and P_{rs} represent the probability of target item identified by a human and by Nao robot. P_{HMS} and P_{rMS} identify the system probability results for humans and a Nao robot, true or false. P_{FHs} and P_{Frs} are the false human and Nao robot probability for correctly rejected. The objective function was calculated for ten numbers of hand objects (*N*). The three parameters *N*, P_{rs}, and P_{Hs} determine both Nao robot's and humans' environmental conditions with values set between 0.1 and 0.9 for each number. The system cost average value was estimated at 4.2 seconds for each number. This high number is due to Nao multisensory and motor limitations.

The first experiment's main task was to enable Nao to interact with children and answer simple mathematical questions using hand gestures and speech. The children would interact with the physical Nao after activating its vision and speech modules to recognize the number of a human agent's fingers in a classroom environment. Figures 4 and 5 show the initiation of Nao agent speech and vision functions.

For the speech, Python code was first compiled using an Anaconda environment, and a file execution ("file01_d.exe") was uploaded to the command NaoQi prompt. The voice module is tested by giving voice commands. Nao's response action is witnessed when getting orders from a human agent, and the command prompt indicates that Nao is ready. Next, a conversation and interaction command is loaded. For example, for the number recognition,

(i) Use a simple pythonic OCR engine using opencv and numpy. http://stackoverflow.com/questions/9413216/simple-digit-recognition-ocr-in-opencv-python

(ii) Run the program using Python example.py

(iii) Press any key a few times for each processing

In the second experiment, the auditory module is activated, and a learning auditory guessing game was used. In this game, a child was asked to calculate the product of two numbers, and the robot reacted by making a clapping sound if the mathematics answer was correct or an alarm if it was not. Next, vision and auditory modules were activated and played interchangeably. This teaching game continued until the child learned from earlier errors, and the Nao recognition system continuously improved as it teaches more children and acquires more data. Thus, an agent would navigate and recall the learned action when an agent interacts with other human agents in the same lab environment [8]. A sample of an output file was generated (as shown in Figure 6).

To measure speaker independence, a kappa ratio measure was implemented to compare agreement of two classifiers having pairs of utterances, one recognized by human,

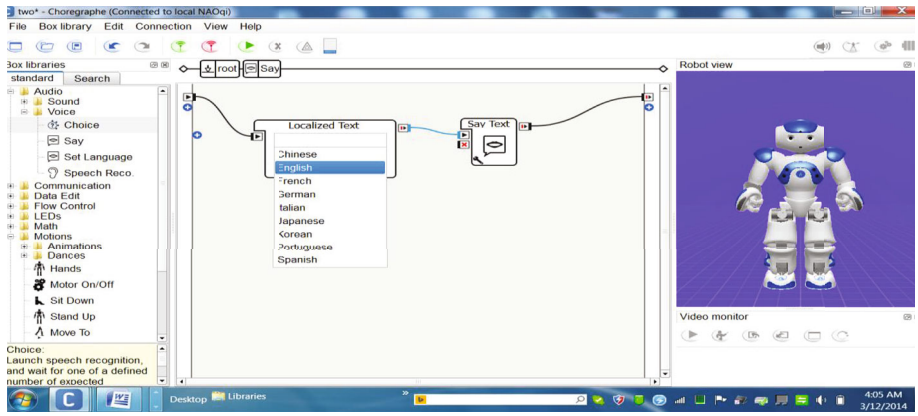


Figure 5: Nao agent speech initiation

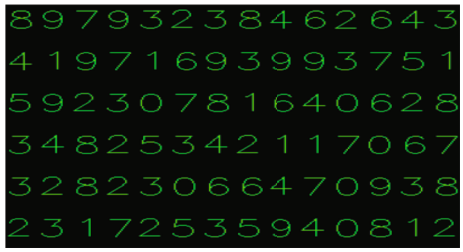


Figure 6: Results of accurately recognized numbers.

Figure 7: Nao participating in the sixth research event at KAU.

while the other by Nao agent, applying the same methodology of the linguistic speech agent of [34] by implementing an HMM syllabic recognizer having a corpus that exceeded 390,000 segment of total pairs. The probability of agreement is equal to 0.50833 measured by

$$\text{Kappa} = \frac{(P(A) - P(E))}{(1 - P(E))}, \tag{3}$$

where $P(A)$ is the actual recognizer and $P(E)$ is the assumed random agreement. The rise of the learning curve for the HMM Nao syllabic recognizer showed a constant logistic growth of 0.030 per iteration with a kappa score of 32.5%, while the human agent scored 40.5% with 0.026 per iteration. For Nao agent, an HMM syllabic recognizer with a kappa ratio 90.8% scored more than 93% of success rate when activating both auditory and vision modules for the agent Nao.

========== HMM Nao Results Analysis ==========

Ref: NaospeechAgent.mlf

Rec: recoutsNaoSpeechAgent.mlf

————————————————————

Overall Results

————————————————————

SENT: %Correct = 95.37 [$H = 103$, $S = 5$, $N = 108$]

WORD: %**Corr** = **94.30**, **Acc** = **93.37** [$H = 104$, $D = 0$, $S = 4$, $I = 1$, $N = 108$]

==

In the third experiment, the author increased the complexity by activating different learning behaviors using several modules within different physical environments, namely, classroom, home, and lab, for 50 hours of training per environment. The module, called ALBehaviorManager, consists of 18 built-in methods. Among these are getInstalledBehaviors(), preloadBehavior(), runBehavior(), addDefaultBehavior(), getUserBehaviorNames(), and isBehaviorPresent(). In the classroom environment, Nao interacted with students and answered their questions using a guessing game. Then, Nao was taken to the second environment at the university laboratory to solve basic math problems. Third, Nao was taken home and to a MOE theater (see Figures 7 and 8) to observe the family members' behaviors by interacting with them. The author implemented the three-constituent principle to focus on Nao's desired interactive behaviors within its morphological capability and limitations. The "here and now" perspective explains Nao's actions, such as answering math questions when asked. The learning and development behavior of Nao is described while answering new questions from previous responses, and evolution is concerned with how Nao's learning behavior evolved by answering unlearned questions or how new behavioral mechanisms emerge. In the "here and now" perspective, the mechanisms and principles are concerned with how behavior comes or how individual behavior results from an agent's interaction with the environments, as explained by Hafner and Möller [37]. The three-time-scale "here and now" led to Nao's instant action corresponding to specific learning-teaching situations and unexpected behavior that enable an emergent

unprogrammed action. Thus, Nao exhibited a defined behavioral relationship with each specified location.

The sensors and controllers developed their perceptual cues and representation models for each environment, and Nao linked specific tasks to each environment using the available module. The author agrees with Pfeifer and Bongard [9] regarding the concept of dynamical systems and attractor states that result from interactions among the input channels (vision, hearing), the ecological niche of different environments (mixed reality, physical, and virtual spaces), and the growth of knowledge from different experiments in understanding (signs, words, and numbers). Once the learning has been completed from a longer time scale, Nao can recognize a new, unprogrammed number and behavior, which indicates the notion of emergence, although this is not the study's focus. For example, when Nao was placed in the home environment observing Muslim prayers, a desired unprogrammed behavior emerged to allow Nao to simulate the human body performing the Muslim prayer actions, such as bending and hand movements. This shows that agent Nao's behaviors can emerge from interactions with the environment. In the final stage of the experiment, the learning was based on mixed reality system environments. Performing teaching-learning tasks over an extended period in different environments caused Nao to become more intelligent and develop a new task in addition to the desired tasks. When Nao is placed in a mixed reality environment, the robot recognized additional gestures, such as body bending and hand movements. The mixed reality environment is used in various fields, such as earth science, engineering, and medicine. They introduce extensive interaction along with a virtual environment to research. According to Schouten et al. [8], any game experience must meet three criteria: context-awareness, adaptation, and personalization. These criteria lead to enhanced interaction from the participant (child). Physical agents in the form of a humanoid robot Nao, with enhanced interaction using a mixed reality system, are making the learning process more interesting and effective than interaction with virtual agent character. For a sample of the time to task interaction and learning between Nao and human agents in different environments, see Table 1. As the number of tasks increases, the Nao agent requires much time to perform the tasks, especially if the interaction level is maximized. The NaoQi Library Python software is activated by calling the built-in modules for the interaction and multiperception recognition time measurements. Thus, the NaoVi module has the fastest reactions (see Table 2).

The interaction between humans and robot agents has been investigated for many years. The use of hardware-based devices and computer vision-related techniques for interaction has also been studied [38]. The choice between hardware- and software-based solutions leads to a tradeoff between accuracy and ease of interaction. In the case of computer vision-based solutions, marker-based and marker-less techniques are used. The use of gloves in marker-based computer vision techniques provides high accuracy but uses a virtual environment. The Haar classifiers technique has been implemented as part of a marker-less computer vision approach. The Haar Algorithm is developed by Viola and Jones [39–44] and involves two main steps (features extractions and objects detections) and is known as "The Haar-training," the normalization threshold in this experiment was set between [1,-1], and the system process can be briefly described in the following steps:

Figure 8: Nao performing at the MOE.

(1) Identify positively detected images patterns

(2) Choose negative images patterns

(3) Select from the positive images the training dataset

(4) Select from the negative and positive images the testing dataset

(5) Employ Haartraining to the selected training dataset

(6) Calculate the performance for the testing dataset

The standard Haar algorithm using OpenCv for figure image processing is summarized below:

Haar uses layers of classifiers; each is trained to detect an object in a specified environment within an image. For each layer, a window is created to match the image and evaluate the information accuracy. If no image matching occurs, the classifier window is said to be "negative." The window matching for the object is reinitiated with another classifier. If the result is "true positive," the matching scores as a successful classification for the positive image. If the result is "false positive," the matching indicates misclassification for negative values as a positive image. If the result returns "false negative," the matching indicates misclassification for positive values as a negative image. The training for the classifier continues until the best score is reached before overtraining is reached. The window passes through all classification layers, with a positive score indicating successful detection [45]. The work shows that the Haar classifier's main disadvantage is that it highly scores false positives in real-time when the number of Haar classifier layers reached 25 if the object keeps moving. The best accuracies of the Haar layers classifier is 23 with 50 training images (see Table 3).

In the fourth environmental experiment, a camera focused on the working space featuring the robot and the child. Information from the robot agent was provided to the child using its motions and projections on a screen. The camera was mounted above the area to provide a top-down view [46]. According to Sugimoto et al. [11], such a view controls the robot from a 2D perspective easier. The presence of

Table 1: Time to task activity Nao-human agents in the four environments.

^{O}RobEnv	^{O}HAgInter	^{O}RobAgInter	^{O}HAgtask	^{O}RobAgtasks	^{O}Ts
1 = class	0	1	1	1	3.3 s
2 = lab	1	0.5	2	2	5.0 s
3 = home	0.5	1	3	3	6.9 s
4 = mixed reality	1	1	4	4	8.6 s

the top-down camera limits the area of operation in the environment. The focus area was fixed, and human-robot interaction only occurred with this specified locality. This limitation did not impair the system because its main objective was to enable the robot to teach the child, who is in its vicinity in any case. The recognition of the child's gesture intended for the robot was extracted using various known techniques using LMC that affected the research for two reasons: the hand's presence in front of the face and the background's lamination. The author used recognition and handheld device based for robot-human interaction to provide haptic feedback from the child to the system [35]. The robot taught the child how to perform mathematical tasks by projecting prerecorded audio or video or gestures. The projection required dimming the lights, whereas the camera focused on the robot, and the child needed proper illumination. Therefore, appropriate lighting or dimming or a handheld projector was necessary [13, 45].

3. Results and Discussion

The proposed system consists of five components: a projector, camera, child, robot, and server (see Figure 9).

(i) Camera: it focuses on the area of coverage. Its position and focus are fixed. The camera's output is regularly passed to the server that uses Haar classifiers to identify the child using face recognition. In the absence of the camera, the robot's eyes are used to input the video, and the face recognition algorithm of the humanoid robot is used to detect the child's presence

(ii) Child: it is the main component of the system, in which learning is the sole objective. To make learning fun for the children, they are given various choices from which to choose, including interaction through a handheld device, speech, fingers of the hand, or LMC

(iii) Robot: it is used to detect the child's finger movements and recognize speech from him/her. The robot can perform face and speech recognition using its default module or pass the acquired data to the server for processing. A communication server runs on the robot to interact with the handheld device of the child

(iv) Server: it is used to control the flow of operations. Instructions from the server can be given either to the robot or the projector. The server can also perform face recognition using a Haar classifier or other means. It can also process the detection of the child's fingers using the convex hull approach and can perform speech recognition

(v) Projector: it is used to display appropriate learning material for the child, which is done only when the child must be taught through already recorded videos. The process can be affected by light in the coverage area; therefore, a suitable projector must be used

Table 2: Multiperception recognition times in the Nao robot.

Library	Hear	See	Speak	Move
NaoQi/actuators	0.857 s	0.231 s	0.252 s	0.752 s

The proposed toolkit description is shown in Figure 10. The process starts with Nao's eye as input using the camera for image processing while the audio is input via Nao's ear. Both inputs will end up with the learning management system. The fixed camera and/or Nao's eye are used to acquire the video of the child. The initial image processing is carried out to check whether the child is in front of the robot. The image acquisition and image processing are described in the flowchart (see Figure 11).

The steps involved in image processing are as follows.

(i) Face identification: the human face in front of the robot is detected using the ALFaceDetection module in Nao. The detected face is written to the ALMemory periodically. Once the Nao robot detects the child's face, it welcomes the child and sends a request to the camera or itself to start acquiring the image

(ii) Restrictive face recognition: the Haar classifiers regulate face recognition on the control server to enable the system to work with only specific children. The data acquired are passed to the control server to recognize already known faces

(iii) Acquisition: vision is implemented through Nao's eyes in the form of an image, and a sequence of images periodically is considered equivalent to a video

(iv) Obtaining image from Nao: the specifications of the image are entered using ALProxy with the ALVideoDevice module. The generic video module (GVM) is used to provide the necessary image format and specifications. The image is obtained using

Input: read image file
Output: number of fingers detected in already detected hand
Step 1: Change the image into grayscale for feature extraction using both cv2.cvtColor () and cv2.inRange ()functions
Locate the hand(s) of the grayscale image and maps onto the colored image and resize using resizing using cv2.getPerspective functions
Match features of the hand(s)/finger(s) segments with the rectangular box by
Detecting pixel coordinates for hand(s) previously saved using cv2.threshold.function
Detecting fingers pixel coordinates inside the hands' coordinates and draw the rectangles
(1) All the number of fingers detected and compared in hands using function cv2.matchTemplate () and calculated using Haar learning algorithm
If (p1, q1) and (p2, q2) are coordinates of the hand and (r1, s1) and (r2, s2) are coordinates of the number of fingers (fingers are present inside the hands)
Then,
p1 < (r1 and r2 both) < p2
q1 < (s1 and s2 both) < q2
Otherwise,
: the number of fingers is Discarded (lie outside the hand).
Step 2: Detection for the number of fingers.
For the remaining fingers, checked existence, add to count for detecting another finger.
If (there is a finger within the defined hand coordinates
Then,
Add one.
Loop: Search for more fingers
For each finger found, find the region of interest of hand which makes the
Square, which represents the finger in the hand
Condition: The number of fingers in one hand should not be greater than 5
the rectangle is extended on both sides
Step last: End.

Algorithm 1

Table 3: Best accuracies of the Haar classifier with 50 training images.

Detected (true positive)	False-positive static objects	False-positive moving objects	Negative (not detected)	False negative	Percentage
47	1	2	1	1	94%

the getImageRemote method and is converted from a pixel image into a PIL image

(v) Conversion: the obtained image is converted into a grayscale or HSV scale. The author used grayscale, where no morphological effects are observed. The HSV scale should be used in case of morphological effects

(vi) Children wearing gloves for LMC are processed using HS. Colored gloves are extracted from the image, thereby separating the fingers of interest from other acquired image/video parts

(vii) Morphological effects: certain morphological effects, such as erosion, dilation, and gradient, are applied to the image/video if the acquired image is different from the user's requirements. This process aims to improve the value of the acquired information

(viii) Blurring: a blur using a Gaussian filter is applied to the gray image to delete the image's Gaussian noise

(ix) Thresholding: the blurred image is then processed in a threshold. This process converts grayscale into a binary image based on the threshold value

(x) Finding contours: contours refer to the outline of the given image. The hierarchy or relationship between contours is obtained, and they are compressed to save space

(xi) Contour areas: the area of each contour, is obtained. The contour with the maximum area is identified and passed to the next stage

(xii) Convex hull and moments: the convex hull is used to find the approximate curve by considering convexity defects. The moments are used to find the center of the given contour. A red circle shown in Figure 12 is drawn at the center of the contour

(xiii) Polygonal curve: the Douglas-Peucker algorithm is used to draw a polygonal curve on the given image. The convex hull is again applied to the output of the polygonal curve

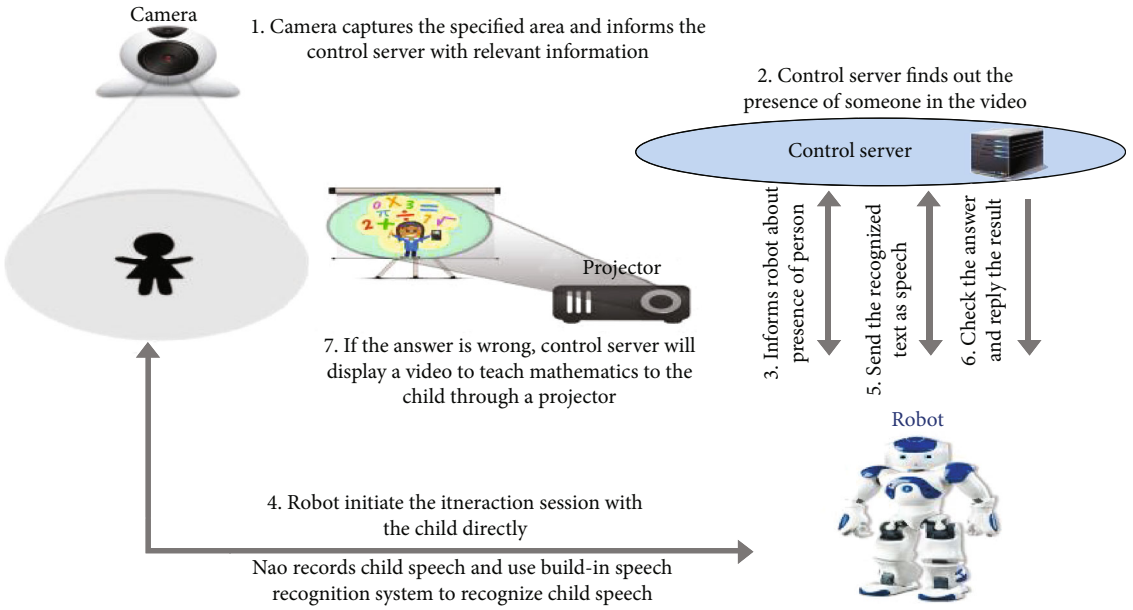


Figure 9: System architecture.

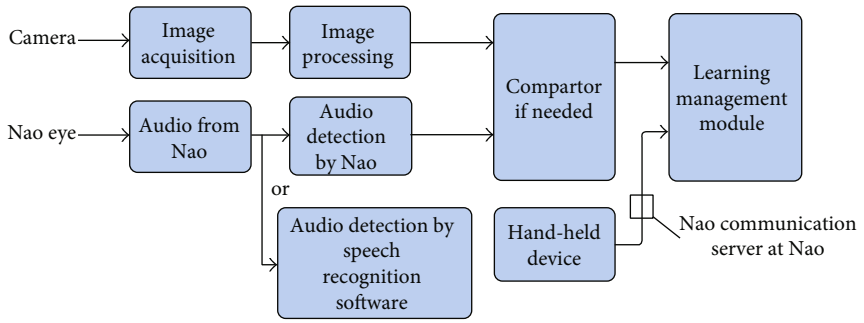


Figure 10: Toolkit description.

(xiv) Convexity defect: any deviation from the convex hull is found with representations as starting and ending points. The appropriate lines and circles are drawn using the result of the convexity defection operation

(xv) Calculating the number of fingers: the points obtained from the convexity defect are used to find the angles between them to decide on the fingers' number

The fingers' number held up by the child must be identified and matched with the trained datasets, the number that the child is showing to the robot Nao due to a simple operation. Various factors hinder the identification of the fingers:

(i) The effect of the cloth (see Figure 12(c))

(ii) The effect of the position of the face (when it is away from the robot) and the fingers (see Figure 12(i))

(iii) The consequence of wearing an ill-fitting glove by the child (see Figure 12(h) and see Figure 11(j))

(iv) The use of colored gloves affecting identification by Nao (see Figure 12(j))

As shown in Figure 12, the system detected the fingers correctly, except for when hindered by the face or other objects. Thus, a glove with an appropriate color was used.

Initially, the sound is detected in Nao using its SoundDetected module. The ALSpeechRecognition module recognizes the voice of the child in Nao. A specified vocabulary dataset (1) list containing one, two, three,….., n, start, and the end is given to the ALS speech recognition module to recognize. Based on the recognizer's confidence level, its efficiency is identified, leading to its acceptance or rejection. As an alternative, speech recognition can be carried out on the server. Streaming audio from Nao is not possible at present; therefore, the stored audio received for a few seconds is transferred to the server for processing. While the server was processing the received audio, Nao continued to record further audio. The storage format used was WAV with four channels at 16 kHz. These channels were used to acquire audio signals from Nao's four receivers, namely, left, right, front, and rear. File transfer was

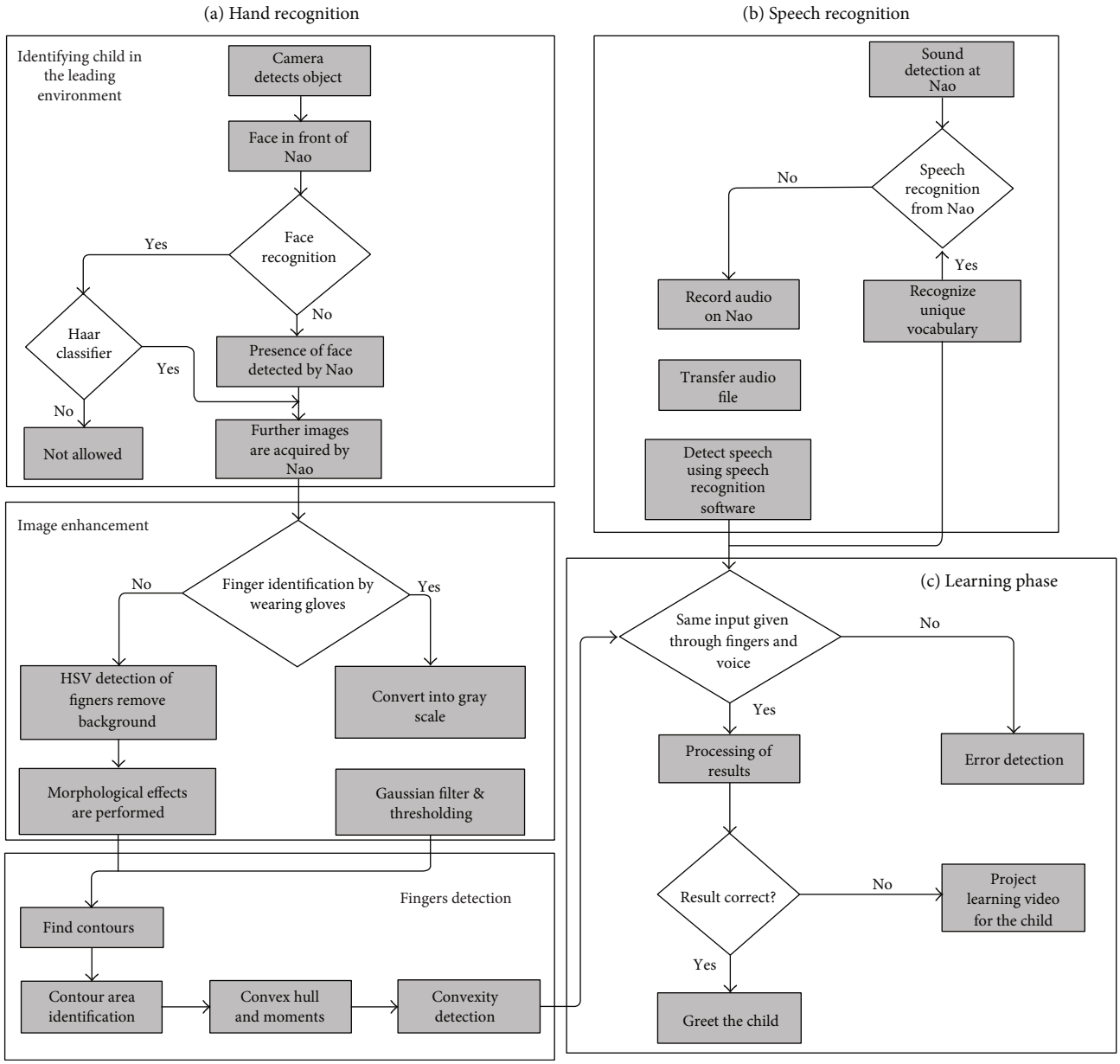


Figure 11: Flowchart of components of the toolkit.

performed using the ALFileManager module. Recognition at the server was carried out using speech recognition software, such as DragonFly. The fixed camera is used to follow the child's location, while Nao initiates simple mathematics teaching for the child. The child's learning is tested based on the output of the learning process. The outcome is displayed on the projected screen by the robot. Thus, both the robot and the projector screen respond to the child. The child also had a mobile device to interact with the robot, and its process of operations is as follows.

The camera recognizes the location of the child.

(i) The child initiates communication with the robot using the mobile application available on his/her mobile device

(ii) The robot teaches a lesson based on tracking audio, video, and movements. Given that Nao has only three fingers, using them to indicate numbers is not feasible

(iii) Nao generates a question for the child to answer

(iv) he response of the child is stored and tested using speech recognition HMM tools to check the answer

(v) On the basis of speech recognition results, an appropriate response is displayed on the projector by the Nao

The system retains ambiguities, as shown in Figure 12, despite the image processing. Changes to the background and luminosity influence the quality of image processing,

Figure 12: Recognition of number of open fingers.

thereby affecting finger recognition. Therefore, a handheld (Android) device-based application was used that communicates with the communication server in Nao. The communication server waits for a connection at its socket [47]. The android application's user interface contains numbers from zero to nine and a few mathematical operations (see Figure 13). The request is passed to the communication server in Nao when the child clicks on the operands, operation, and result. The server checks the result. If it is correct, Nao greets the child with a clapping sound and hand movement action. If not, a video on the specific operation is projected for the child from the control server. Options are provided for the child or a parent to change the robot's voice volume and speed.

The mixed reality system uses different entities to enhance learning. The performance of the system when using

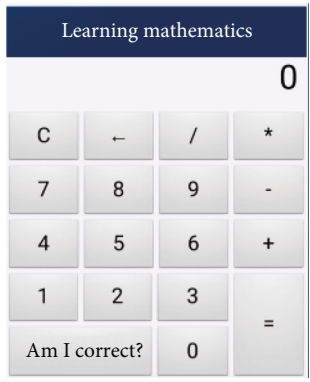


Figure 13: Initial screen of the handheld application.

Table 4: Performance comparison.

Criteria	Measure
Communication distance between Nao and human	0.77 m to 1.12 m
Best communication distance between Nao and human	0.9 m

the Haar classifier and Nao's facial recognition algorithm was compared. Both correctly identified the face in front of the robot. Nao's facial recognition drawback was that it expected the face to be close and needed some time to recognize the face. According to documentation for the ALVisionRecognition-Nao Software 1.14.5 [48], the recognition process works between half and twice the distance used for learning purposes, achieving a 95% success rate, while the LMC using Haar classification scored 98.50%. The closeness of the human face to Nao restricted the child's free movements indicated in Table 4.

The use of different image processing techniques to recognize the number of fingers was significantly affected by the position of the face, the background, and the area's luminosity [49–53]. Therefore, a handheld device-based interaction system was provided for the child to interact with the robot. The device was connected via WiFi to the robot to reduce the restriction of a child's mobility. The projector in this mixed reality system also improved the child's learning capability due to excitement in dealing with Nao. The basic limitations of the study are

(i) Children should be familiar with Nao using Arabic number language to communicate in gestures

(ii) The fast movement speed affect LMC and recorded as errors

(iii) Nao and the participants' distance should not exceed 0.09 m due to Nao's hardware audio-visual limitations

(iv) The data sets size and quality could be improved for robust real-time recognition

(v) Setting up the proper illumination level is necessary

(vi) Streaming multisensory perception cues from Nao is not possible at present

(vii) Nao has only three fingers, not five, to communicate the number visually with the child

4. Conclusions and Future Work

In this work, the author developed a toolkit and evaluated the results in a mixed reality environment to enhanced learning by children and increased the robot Nao's intelligence level from a 3- to 7-year-old child. The teacher, in this case, is the robot Nao who interacted with the child through various means and environments. Four experiments were conducted to test interaction in four different environments (class, lab, home, and mixed reality using Leap Motion Controller). The author showed that implementing an AI principle design, namely, the three-constituent principle, helped grow the robot's intelligence using different environments. The developed toolkit, using Arabic speech recognition and the Haar algorithm for robust image recognition in a mixed reality system architecture implementing big data, enabled Nao to gradually achieve a higher learning success rate ranged from 90.8%, 93%, 94%, to 98.50% as environment changes and multisensory perception increases. The highest learning level was achieved using LMC hand sign gestures with the Haar algorithm featuring a mixed reality environment. Activating a multisensory perception of vision, hearing, speech, and gestures for Human-Robot Interactions (HRI) in real-time increases children's learning math experience and makes it more enjoyable. The implementation of Arabic Speech Agent for Nao using phonological knowledge activated for HRI communication. The study shows that Nao's robot intelligence could be increased by learning similar to human intelligence and teaching simple mathematics to children. A cutting-edge research work direction for fostering Child-Robots education could be achieved using an active warehouse multidata system. Improving the Haar algorithm to operate in real-time with multihuman agent interaction and a single robot is one step toward future learning. An ultimate digital or physical robot teacher that operates in real time could be set as a goal for the years to come.

Data Availability

Preliminary training datasets for Nao involved: https://github.com/IBM/watson-nao-robot, Andreas Ess' webpage (ethz.ch), The NAL dataset (inria.fr), SpeechRecognition · PyPI, Baothman speech corpus [49], Leap Motion Controller [19].

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The author thanks the Science and Technology Unit at King Abdulaziz University. This project was funded by the National Plan for Science, Technology, and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, the Kingdom of Saudi Arabia (award number 03-INF188-08).

References

[1] D. Poole, "The independent choice logic for modelling multiple agents under uncertainty," *Artificial Intelligence*, vol. 94, no. 1-2, pp. 7–56, 1997.

[2] R. Solomono, "The time scale of artificial intelligence: reflections on social effects," *Human Systems Management*, vol. 5, no. 2, pp. 149–153, 1985.

[3] H. S. Nwana, "Software agents: an overview," *Knowledge Engineering Review*, vol. 11, no. 3, pp. 205–244, 1996.

[4] W. Brenner, R. Zarnekow, and H. Wittig, "Application areas for intelligent software agents," in *Intelligent Software Agents*, Springer, Berlin, Heidelberg, 1998.

[5] W. Shen and D. H. Norrie, "Agent-based systems for intelligent manufacturing: a state-of-the-art survey," *Knowledge and Information Systems*, vol. 1, no. 2, pp. 129–156, 1999.

[6] J. Tweedale, N. Ichalkaranje, I. Agents, and T. Applications, "Intelligent agents and their applications," in *Knowledge-Based Intelligent Information and Engineering Systems. KES 2006. Lecture Notes in Computer Science, vol 4252*Springer, Berlin, Heidelberg.

[7] V. Kumar, A. Dixit, R. G. Javalgi, and M. Dass, "Research framework, strategies, and applications of intelligent agent technologies (IATs) in marketing," *Journal of the Academy of Marketing Science*, vol. 44, no. 1, pp. 24–45, 2016.

[8] B. A. Schouten, R. Tieben, A. van de Ven, and D. W. Schouten, "Human behavior analysis in ambient gaming and playful interaction," in *Computer Analysis of Human Behavior*, A. Salah and T. Gevers, Eds., pp. 387–403, Springer, London, 2011.

[9] R. Pfeifer, F. Iida, and J. Bongard, "New robotics: design principles for intelligent systems," *Artificial Life*, vol. 11, no. 1-2, pp. 99–120, 2005.

[10] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi, "TouchMe: an augmented reality-based remote robot manipulation," *the 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011, 2011*, 2011http://www.ic-at.org/ICAT2011_Proceedings/pdf/061-Hashimoto.pdf.

[11] M. Sugimoto, G. Kagotani, H. Nii, N. Shiroma, M. Inami, and F. Matsuno, "Time follower's vision," *ACM SIGGRAPH Emerging technologies*, p. , 200429, 2004, http://dl.acm.org/citation.cfm?id=1186185.

[12] F. Leutert, C. Herrmannand, and K. Schilling, "A spatial augmented reality system for intuitive display of robotic data," *Proceedings of the 8th ACM/IEEE International Conference on Human–Robot Interaction*, pp. , 2013179-180, 2013, http://dl.acm.org/citation.cfm?id=2447626.

[13] A. Moreno, A. van Delden, R. Poppe, and D. Reidsma, "Socially aware interactive playgrounds," *IEEE Pervasive Computing*, vol. 12, no. 3, pp. 40–47, 2013.

[14] I. F. Siddiqui, N. M. F. Qureshi, B. S. Chowdhry, and M. A. Uqaili, "Edge-node-aware adaptive data processing framework for smart grid," *Wireless Personal Communications*, vol. 106, no. 1, pp. 179–189, 2019.

[15] N. M. F. Qureshi, I. F. Siddiqui, M. A. Unar et al., "An aggregate mapreduce data block placement strategy for wireless IoT edge nodes in smart grid," *Wireless Personal Communications*, vol. 106, no. 4, pp. 2225–2236, 2019.

[16] J. Koo and N. M. F. Qureshi, "Fine-grained data processing framework for heterogeneous IoT devices in sub-aquatic edge computing environment," *Wireless Personal Communications*, vol. 116, no. 2, pp. 1407–1422, 2021.

[17] I. F. Siddiqui, N. M. F. Qureshi, B. S. Chowdhry, and M. A. Uqaili, "Pseudo-cache-based IoT small files management framework in HDFS cluster," *Wireless Personal Communications*, vol. 113, no. 3, pp. 1495–1522, 2020.

[18] A. K. Podder, A. Al Bukhari, S. Islam et al., "IoT based smart agrotech system for verification of urban farming parameters," *Microprocessors and Microsystems*, vol. 82, article 104025, 2021.

[19] M. K. A. GHANI, M. A. Mohammed, M. S. Ibrahim, S. A. Mostafa, and D. A. IBRAHIM, "Implementing an efficient expert system for services center management by fuzzy logic controller," *Journal of Theoretical & Applied Information Technology*, vol. 95, no. 13, 2017.

[20] M. A. Mohammed, B. Al-Khateeb, and D. A. Ibrahim, "Case based reasoning shell frameworkas decision support tool," *Indian Journal of Science and Technology*, vol. 9, no. 42, pp. 1–8, 2016.

[21] S. A. Mostafa, A. Mustapha, S. S. Gunasekaran et al., "An agent architecture for autonomous UAV flight control in object classification and recognition missions," *Soft Computing*, 2021.

[22] A. Lakhan, Q.-U.-A. Mastoi, M. Elhoseny, M. S. Memon, and M. A. Mohammed, "Deep neural network-based application partitioning and scheduling for hospitals and medical enterprises using IoT assisted mobile fog cloud," *Enterprise Information Systems*, pp. 1–23, 2021.

[23] K. H. Abdulkareem, M. A. Mohammed, A. Salim et al., "Realizing an effective COVID-19 diagnosis system based on machine learning and IOT in smart hospital environment," *IEEE Internet of Things Journal*, 2021.

[24] M. A. Mohammed, S. S. Gunasekaran, S. A. Mostafa, A. Mustafa, and M. K. A. Ghani, "Implementing an agent-based multi-natural language anti-spam model," in *2018 International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR)*, pp. 1–5, Putrajaya, Malaysia, August 2018.

[25] K. H. Abdulkareem, M. A. Mohammed, S. S. Gunasekaran et al., "A review of fog computing and machine learning: concepts, applications, challenges, and open issues," *IEEE Access*, vol. 7, pp. 153123–153140, 2019.

[26] S. A. Mostafa, A. Mustapha, A. A. Hazeem, S. H. Khaleefah, and M. A. Mohammed, "An agent-based inference engine for efficient and reliable automated car failure diagnosis assistance," *IEEE Access*, vol. 6, pp. 8322–8331, 2018.

[27] A. A. Mutlag, M. Khanapi Abd Ghani, M. A. Mohammed et al., "MAFC: multi-agent fog computing model for healthcare critical tasks management," *Sensors*, vol. 20, no. 7, article 1853, 2020.

[28] V. Lahoura, H. Singh, A. Aggarwal et al., "Cloud computing-based framework for breast cancer diagnosis using extreme learning machine," *Diagnostics*, vol. 11, no. 2, p. 241, 2021.

[29] S. N. Makhadmeh, M. A. Al-Betar, Z. A. A. Alyasseri et al., "Smart home battery for the multi-objective power scheduling problem in a smart home using grey wolf optimizer," *Electronics*, vol. 10, no. 4, p. 447, 2021.

[30] S. A. Mostafa, S. S. Gunasekaran, A. Mustapha, M. A. Mohammed, and W. M. Abduallah, "Modelling an adjustable autonomous multi-agent Internet of Things system for elderly smart home," in *Advances in Neuroergonomics and Cognitive Engineering. AHFE 2019. Advances in Intelligent Systems and Computing, vol 953*, H. Ayaz, Ed., Springer, Cham, 2020.

[31] N. Bostrom, "Are we living in a computer simulation?," *The Philosophical Quarterly*, vol. 53, no. 211, pp. 243–255, 2003.

[32] G. Fisher, "User modeling in human-computer interaction," *User Modeling and User-Adapted Interaction*, vol. 11, no. 1/2, pp. 65–86, 2001.

[33] R. Pfeifer and J. Bongard, *How the Body Shapes the Way We Think—A New View of Intelligence: A Bradford Book*, The MIT Press, Cambridge, Massachusetts, 2006.

[34] F. Baotrhman, *Phonology Based AutomaticSpeech Recognition for Arabic*, PhD thesis, The University of Huddersfield, 2003.

[35] A. A. Almarzuqi, *Gesture-Based Assistive Robotics Children Education through Enhanced Interaction*, MSc. Thesis, King Abdulaziz University, 2017.

[36] D. Yashpe, *Influence of Human Reaction Time in Human-Robot Collaborative Target Recognition Systems*, M.Sc thesis, University of the Negev, 2009.

[37] V. V. Hafner and R. Möller, "Learning of visual navigation strategies," *Proceedings of the European Workshop on Learning Robots*, vol. 1, pp. 47–56, 2001.

[38] A. B. Raij, K. Johnsenand, and D. S. Lind, "Comparing interpersonal interactions with a virtual human to those with a real human," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 3, pp. 443–457, 2007.

[39] A. Kumar, M. S. Hashmi, A. Q. Ansari, and S. Arzykulov, *Haar Algorithm for The Analysis of Fractional Order Calculus Based Computation Problems Associated With Electromagnetic Waves*, Research Square Co.,Inc, 2021.

[40] P. Taunk, G. Jayasri, J. P. Priya, and N. S. Kumar, "Face detection using Viola Jones with Haar cascade," *Test Engineering and Management*, vol. 83, 2020.

[41] Y. Feng, Q. Jia, and W. Wei, "A control architecture of robot-assisted intervention for children with autism spectrum disorders," *Journal of Robotics*, vol. 2018, Article ID 3246708, 12 pages, 2018.

[42] J. Huang, Y. Shang, and H. Chen, "Improved Viola-Jones face detection algorithm based on HoloLens," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, 2019.

[43] F. S. Daniş, T. Meriçli, Ç. Meriçli, and H. L. Akın, "Robot detection with a cascade of boosted classifiers based on Haar-like features," in *RoboCup 2010: Robot Soccer World Cup XIV. RoboCup 2010. Lecture Notes in Computer Science, vol 6556*, J. Ruiz-del-Solar, E. Chown, and P. G. Plöger, Eds., pp. 409–417, Springer, Berlin, Heidelberg, 2011.

[44] R. Ogla, A. Ogla, A. Abdul Hussien, and M. Mahmood, "Face detection by using OpenCV's Viola-Jones Algorithm based on coding eyes," *Iraqi Journal of Science*, vol. 58, pp. 735–745, 2017.

[45] *Microvision Display Technology Solutions. n.d.*August 2014, http://www.microvision.com/solutions/.

[46] C. Guo, J. E. Young, and E. Sharlin, "Touch and toys: new techniques for interaction with a remote group of robots," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, pp. 491–500, New York, NY, USA, 2009.

[47] *Nao Communication Server. n.d. northernstars-wiki.wikidot.-com/projects:naocom*http://northernstars-wiki.wikidot.com/projects:naocom.

[48] *ALVision Recognition-Nao Software 1.14.5 documentation. n.d.*March 2015, http://doc.aldebaran.com/1-14/naoqi/vision/alvisionrecognition.html.

[49] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa et al., "Voice pathology detection and classification using convolutional neural network model," *Applied Sciences*, vol. 10, no. 11, article 3723, 2020.

[50] S. A. Kashinath, S. A. Mostafa, A. Mustapha et al., "Review of data fusion methods for real-time and multi-sensor traffic flow analysis," *IEEE Access*, vol. 9, pp. 51258–51276, 2021.

[51] X. Zhou, Y. Ma, Q. Zhang, M. A. Mohammed, and R. Damaševičius, "A reversible watermarking system for medical color images: balancing capacity, imperceptibility, and robustness," *Electronics*, vol. 10, no. 9, article 1024, 2021.

[52] O. A. Mahdi, Y. R. B. Al-Mayouf, A. B. Ghazi, A. W. A. Wahab, and M. Y. I. B. Idris, "An energy-aware and load-balancing routing scheme for wireless sensor networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 3, pp. 1312–1319, 2018.

[53] M. S. P. Subathra, M. A. Mohammed, M. S. Maashi, B. Garcia-Zapirain, N. J. Sairamya, and S. T. George, "Detection of focal and non-focal electroencephalogram signals using fast Walsh-Hadamard transform and artificial neural network," *Sensors*, vol. 20, no. 17, article 4952, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5533161, 14 pages
https://doi.org/10.1155/2021/5533161

Research Article

IoT-Based Smart Management of Healthcare Services in Hospital Buildings during COVID-19 and Future Pandemics

Omid Akbarzadeh,[1] **Mehrshid Baradaran**,[2] **and Mohammad R. Khosravi**[3]

[1]*Department of Electrical and Computer Engineering, Shiraz University, Iran*
[2]*Department of Computer Science and Engineering, SBU Division of Computer Engineering, Iran*
[3]*Department of Computer Engineering, Persian Gulf University, Iran*

Correspondence should be addressed to Omid Akbarzadeh; omidakbarzadeh82@gmail.com

Received 8 February 2021; Revised 4 May 2021; Accepted 18 June 2021; Published 2 July 2021

Academic Editor: Nawab Muhammad Faseeh Qureshi

The paper aims to design and develop an innovative solution in the Smart Building context that increases guests' hospitality level during the COVID-19 and future pandemics in locations like hotels, conference locations, campuses, and hospitals. The solution supports features intending to control the number of occupants by online appointments, smart navigation, and queue management in the building through mobile phones and navigation to the desired location by highlighting interests and facilities. Moreover, checking the space occupancy, and automatic adjustment of the environmental features are the abilities that can be added to the proposed design in the future development. The proposed solution can address all mentioned issues regarding the smart building by integrating and utilizing various data sources collected by the internet of things (IoT) sensors. Then, storing and processing collected data in servers and finally sending the desired information to the end-users. Consequently, through the integration of multiple IoT technologies, a unique platform with minimal hardware usage and maximum adaptability for smart management of general and healthcare services in hospital buildings will be created.

1. Introduction

During the COVID-19 pandemic, high-tech solutions like the Internet of Things (IoT) for smart buildings have been critical in keeping our urban societies functional [1–2]. The aim of highlighting the smart building with IoT technologies is to identify COVID-19 cases, decrease the spread, and reduce the impact of the pandemic. Smart buildings, supported by IoT, were primarily relied upon for security, automated management and control, increasing energy efficiency, safety, usability, and accessibility. Furthermore, as lockdown eases, they will also help manage building occupancy levels and social distancing [3–5].

This research is aimed to improve smart buildings' features by adding queue management, smart navigation of places to minimize people connection, and social distancing safety mechanisms. Besides, environmental advancement features to increase people's safety are features that this project can offer in the subsequent development [6–8].

This solution implements through several steps can be seen in Figure 1. Our model-building to develop our proposed solution is a hospital, which is the most demanded place to make it smart based on surveys [9–12]. As the first step, sensors' positioning is deployed using our Building Information Modeling (BIM) and mathematical formulation. Next, the data will send to the gateway through several specific protocols. Then, received data will be stored in the servers and processed through the designed middleware; consequently, the processed data sent to the mobile application, shown in Figure 2 [13–16]. The presented platform's primary thinking is to propose an application that users can control through their smartphones using the data gathered through sensors and beacons located in the different parts of a hospital [17–20]. IoT would be the best solution to

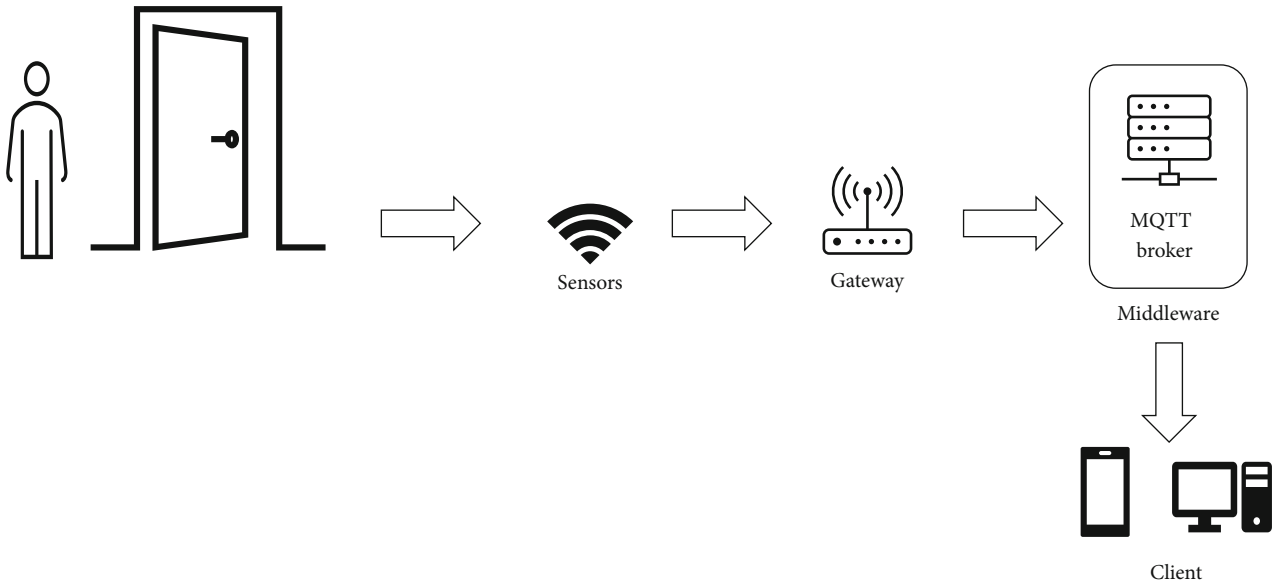


Figure 1: A workflow depicts the general steps and components required for the implementation of the project.

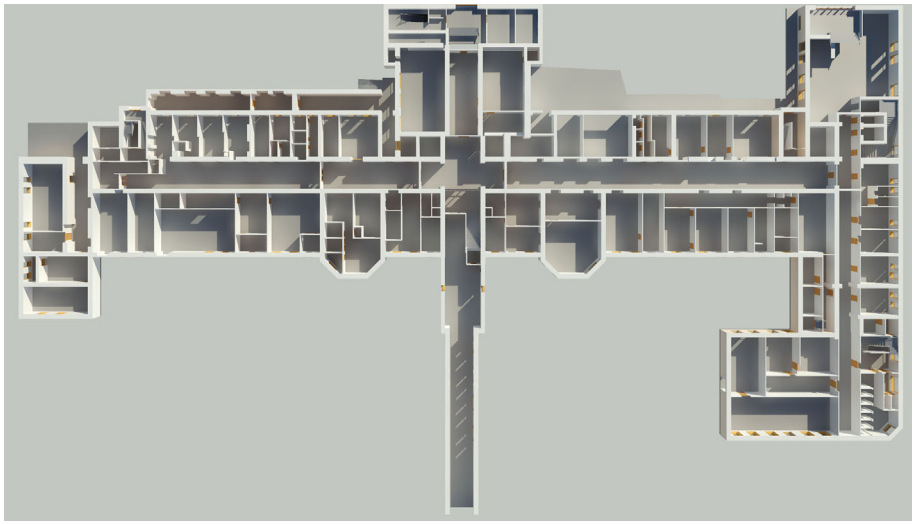


Figure 2: The image shows the BIM model of the building used for deploying sensors.

combine all collected information and send it to the end-users [20].

Since position determining is considered a significant section of these solutions, we have intensely focused on position determining. Previous efforts with indoor positioning systems focus on statistical fingerprinting methods, mainly using 802.11 (WLAN) as the platform. Some efforts were made with purely signal strength-based positioning, but indoor environments have shown to work unfavorably for these kinds of methods [21–23]. The novelty of our presented platform is to take advantage of previous efforts and customize and optimize them to grant them the ability to implement in an actual project since most of the earlier efforts were limited to computer simulation. Moreover, regarding the previous efforts, developing a mobile application that offers the mentioned services to the end-users can be considered a unique feature of the proposed platform [1, 24–30].

One of the major functionalities of the smart building is in healthcare, as a smart hospital. Due to the COVID-19 pandemic, the hospital's occupancy level and social distancing are critical to reducing the communication between patients and medical staff to reduce the spread of COVID-19 pandemic cases [31–37]. Moreover, in the COVID-19 pandemic and other viral illnesses, taking good patient care is practical [38–42].

Our proposed design's supremacy provides a wide range of practical information for end-users, such as a list of exams and queuing guidance to the next room and direction to the desired destinations [43–46]. Furthermore, this scheme provides information about the point of interest, like the number of people, waited in line for service and accessibility of that service at that time [47–52]. Besides, this scheme notifies end-users about breaking the social distance in the situation of the COVID-19 pandemic.

Table 1: Small overview of positioning features technologies using wireless technology.

Wireless technology	Power efficiency	Application (social/industrial)	Positioning accuracy	Advantage	Disadvantage	Some related resources
Our proposed architecture	High	Social (health, commercial, etc.)	High	Low-price	n/a	—
Ultrawideband (UWB)	Low (relatively)	General	High	Easiness for implementation	Technology availability	#ref [5-8]
GPS and similar satellite-based models	Low (relatively)	General	Average	Easiness for implementation	High-price	#ref [5-8]
802.11 fingerprinting	Low (relatively)	General	High	Technology availability	Difficulty for implementation	#ref [5-8]
802.11 time differential lateration	Low (relatively)	General	Average	Technology availability	Difficulty for implementation	#ref [5-8]
Bluetooth fingerprinting	High	General	Average	Low-price	Difficulty for implementation	#ref [5-8]
Cellular proximity	Average (relatively)	General	Very low	Technology availability	High-price	#ref [5-8]

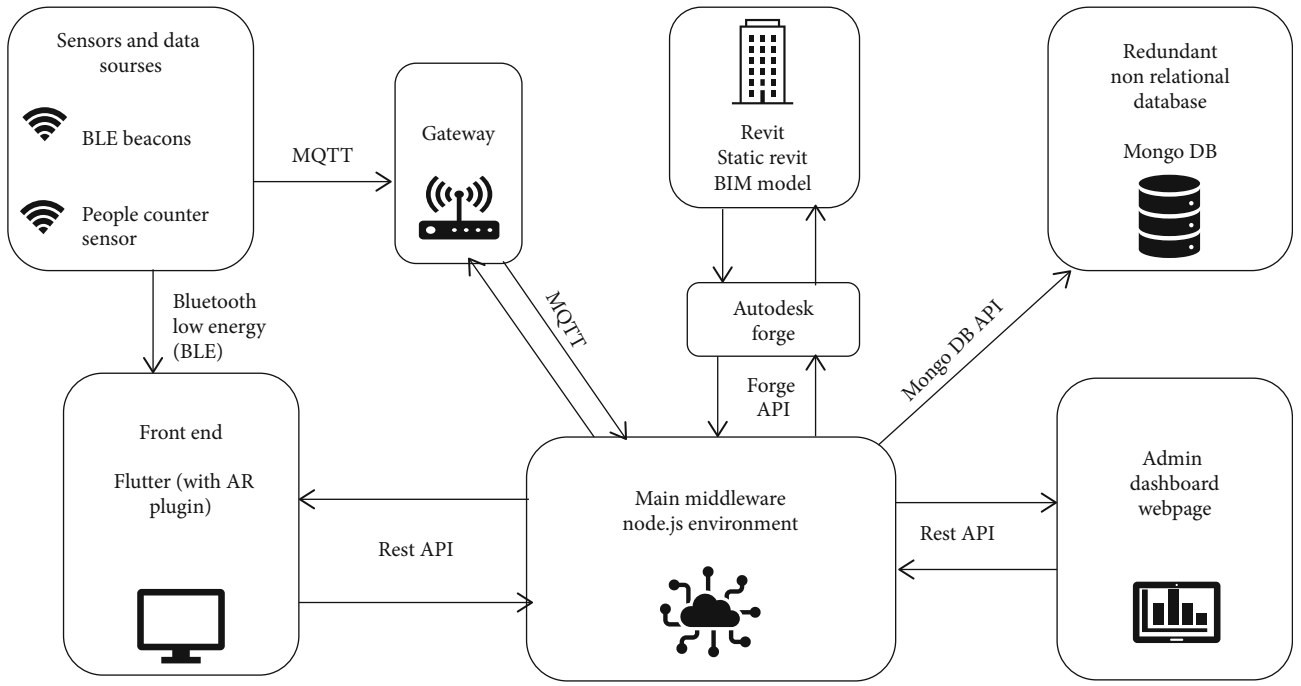


Figure 3: This workflow shows the connection between the different components of the system as well as using technologies.

In Section 3, we discuss the methodologies used in the proposed technique. Furthermore, in this section, we discuss the positioning and related components such as Bluetooth low energy, beacons, and concepts to provide a detailed project definition. Then, in Section 4, we discuss the information collected through surveys and their application to show the mobile application's primary outcome, particularly for the COVID-19 case [53–57].

2. Related Work

The modern technologies used for indoor positioning are intensively various in terms of accuracy, cost, and maintenance requirements, e.g., most of them have concentrated on IEEE 802.11, known as Wi-Fi. Even though the previous works in this field have achieved acceptable results, they have not used power-efficient wireless technology, such as 802.11. But regarding Bluetooth, the initial attempts to use this technology for postponing purposes do not show promising results [5]. Since many parts of earlier versions of Bluetooth standard are not suitable for positioning. The two main problems of positioning technologies, alongside energy efficiency, are the clock accuracy and the received signal strength

Table 2: The table defines the constants used in Eq. (2) empirically.

Constant	A	β	γ
Value	0.889	7.7895	0.111

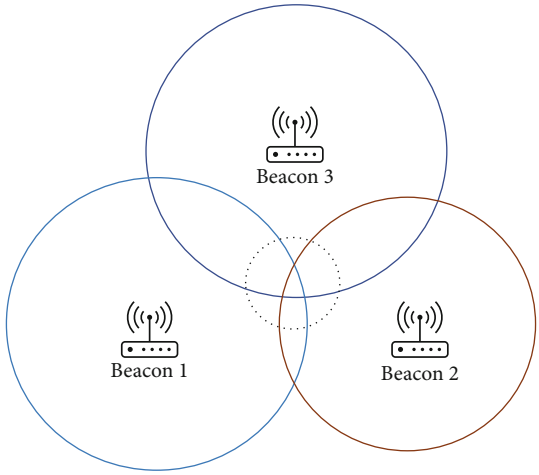


Figure 4: Trilateration-based positioning system [1].

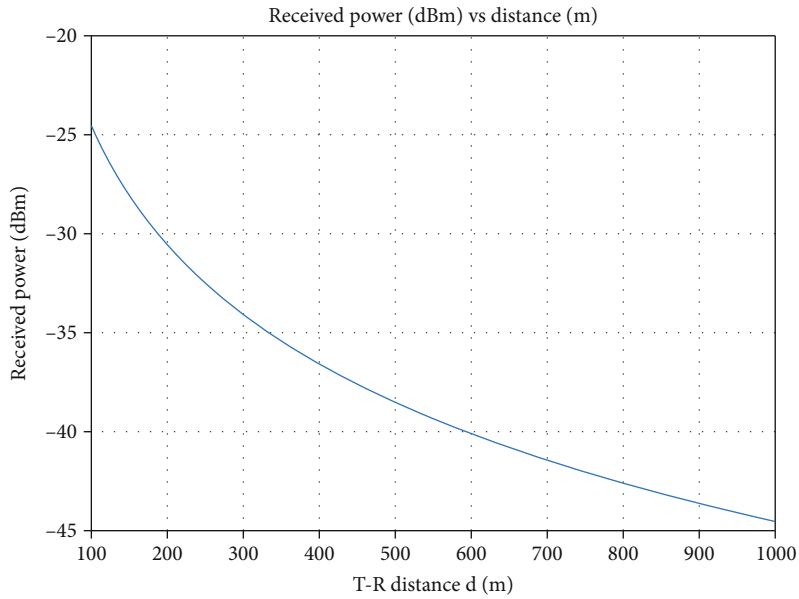


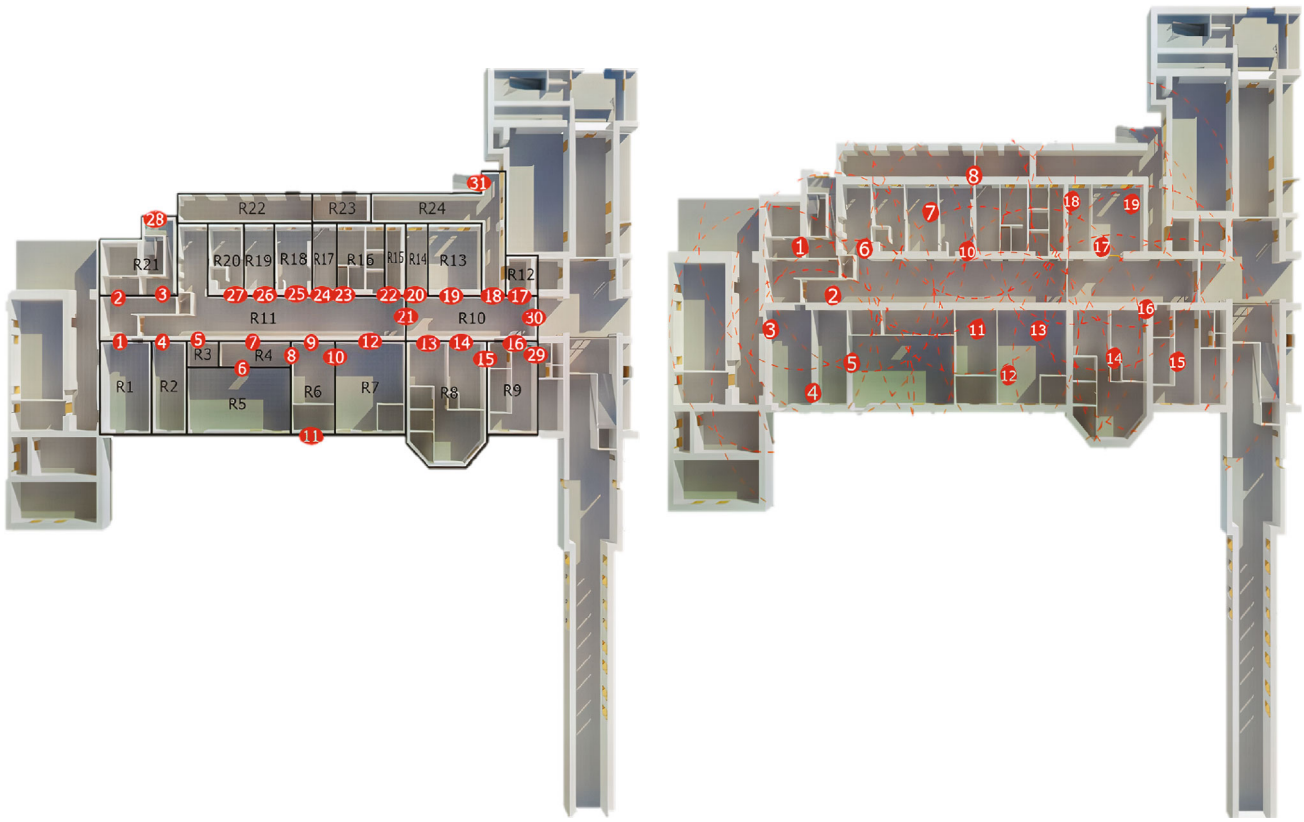
Figure 5: A curve that shows the relation between signal strength at the receiver and distance could be viewed by increasing the distance signal power decreases exponentially.

indication (RSSI) value that affect the expected accuracy level. Unfortunately, we have not found any closely related articles in time-based positioning with a dedicated focus on this.

On the other hand, related work focused on the correlation between distance and RSSI [6]. Although this correlation is practical, the accuracy level is limited in this case. The concept of fingerprinting is a subject of other studies focused on positioning technologies as well. In the Bluetooth-based systems, all the desired location is covered with BLE beacons that constantly propagate their presence. Then, the device planned to be located computes all the received propagations' signal strength and compares those and a precompiled database of locations. Consequently, selecting the best match can achieve an accuracy level of fewer than 3 meters with a very high level of precision [7]. Some papers also focus on Bluetooth, old versions, not BLE, which might have different properties making it more useful for other methods such as RSSI-based ones.

There is almost no clear standard to address several existed commercial indoor positioning systems on the market. In fact, the technology providers do not specifically describe their products. Moreover, the used method and expected level of accuracy are not determined. However, most of them rely on the 802.11 standard, which makes them inherently power inefficient. They are typically used for positioning people, either for navigation or location-specific information such as commercial offers at retail locations or tour information.

Table 1 shows a selection of positioning techniques available for indoor use. Regarding GPS as a reference, the numbers are estimated, and both power consumption and accuracy may vary in the systems based on similar designs. The power figures are measured in the active condition,



(a) (b)

Figure 6: Part (a) depicts the position of people counter sensors, and part (b) shows the positions of BLE sensors in the building.

and it is not directly comparable between each other. Some technologies are energy efficient (drastically less than technology with similar active state power, but setup times are longer).

Therefore, the absence of technology with less than ten meters of accuracy and an acceptable level of power efficiency can be considered the main issue. UWB and BLE-based systems are considered two options that look particularly promising, and both can be used for indoor positioning. Cost and accuracy expectations are still crucial for different technologies and critical while selecting a wireless technology. This paper focuses on BLE-based systems since BLE is deployed in mobile phones and wearable electronics, and it is expected to grow rapidly in its usage. BLE-based systems provide many economic advantages such as low-cost, low-power, and high availability of devices. Moreover, the BLE-based systems would compete with UWB for accuracy and resolution, but as mentioned, lower prices and more availability of technology [8].

3. Proposed Method

3.1. General Architectures. This section talks about the methodology used by the proposed system presents the block diagram of the working of the system at the user end Figure 3. Regarding the provided graph, Figure 3, counter sensors and Bluetooth low energy (BLE) beacons are considered the source of data that sends their data to the gateway through

```
S1room1 = s1
S1room2 = s4
S1room3 = s5
S1room4 = s7-s6-s8
S1room5 = s6
S1room6 = s8+s9-s10-s11
S1room8 = s13+s14-s15
S1room9 = s15+s16-sb
```

Figure 7: The procedure used to implement the people counting sensor.

Message Queuing Telemetry Transport (MQTT) messages. The main idea is to propose the people counting with minimum error using three sensors in each door. These three sensors should work concurrently and calibrate in a single code. The gateway can implement each sensor data using wired or wireless communication. The working principle is a simple IN-OUT principle. Whenever a person goes through the inside (dedicated direction for every room), the optimized data will increase the room's total number. Otherwise, it will decrease. According to the provided workflow, we have a BIM model that includes information about the geometry and generally considered a database.

Furthermore, communication between different blocks of the mentioned workflow performs through MQTT, REST API, and Bluetooth protocols. Admin DASHBOARD is a tool to help the end-user to control the data. Regarding the

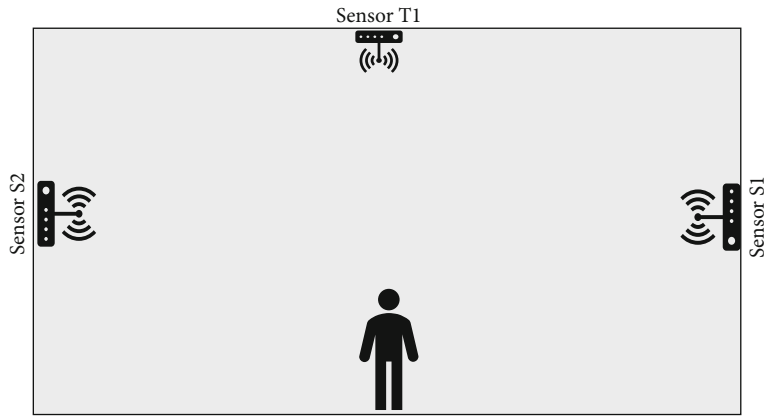


Figure 8: image shows the simple IN-OUT principle.

database, Mongo D.B. and Studio3T are used for this purpose; also, Thingsboard.io is used as the control dashboard.

3.2. Positioning. To discuss the concept of positioning, a brief introduction of basic concepts is necessary, mentioned in the following sections.

3.2.1. Proximity. One of the easiest positioning methods is to check that the object is present in radio coverage without considering the received signal strength indicator (RSSI) or arrival time. The resolution offered is the same as the radio coverage, as it would only differentiate between two states, present and not present. Accuracy will increase through combining several radios at different positions and taking advantage of their coverage areas overlap to determine the object's precise situation.

3.2.2. Ranging Concept. The procedure of distance determination between two objects, without considering the angle. It is also the basic concept utilized for lateration. To accomplish this purpose, several methods were studied. The two most popular approaches in radio ranging are according to received signal strength or time of flight. The formulation between signal strength and distance is expressed as "$d = 20^{(}\mathrm{RSSI} + \alpha)$" in the free space. In the previous formula, α is an offset according to the maximum received signal strength of the system. However, this formulation is an inaccurate estimation for indoor positioning use cases since indoor areas contain many obstacles such as walls and furniture. It is worth noting that the obstacles are the primary reasons for multipath effects that lead to the complication of the relation between distance and RSSI and reduce accuracy.

3.2.3. Lateration Concept. Using multiple range measurements determines one object's position. When the range is measured from several certain positions, each measurement is considered a circle or sphere of possible positions at that range. Combining several measurements and calculating the junction of the measurements lead to the estimation of position. For example, the mentioned procedure is depicted using three beacons located at a different distance from the object.

```
/*subscribe to MQTT broker */
mqtt_subscribe(&client, topic, 0);
/* start publishing the time */
printf("%s listening for '%s' messages.\n", argv[0], topic);
printf("Press CTRL-D to exit.\n\n");
/* block */
while(fgetc(stdin) != EOF);
/* disconnect */
printf("\n%s disconnecting from %s\n", argv[0], addr);
sleep(1);
/* exit */
exit_example(EXIT_SUCCESS, sockfd, &client_daemon);
```

Algorithm 1: It shows the main process.

3.3. Bluetooth: Low-Power Tool. BLE is a low-power radio standard Introduced in the Bluetooth 4.0 standard. Keeping Bluetooth as part of its name does not mean complete backward compatibility with the earlier Bluetooth versions. Bluetooth 4.0 standard specifies two different kinds of radio interfaces, first, compatibility with the previous standard, known as Basic Data Rate (B.R.) or Enhanced Data Rate (EDR). Second, the dual-mode and single-mode can support BR/EDR or BLE, provided through Bluetooth 4.0. The directionality of the antenna is a primary consideration that should be considered with all BLE-enabled devices. Using small antennas leads to the property of nonisotropic radiation patterns. Therefore, in most positioning scenarios, due to antenna directionality and the size of obstacles, adding an external antenna to these devices cannot be considered as a solution. This problem can be addressed using relative signal strengths instead of absolute ones when position determination performs through multiple signals.

3.3.1. Radio Interface. The radio interface is one of the drastic changes regarding the earlier Bluetooth versions. BLE takes advantage of a 2.4 GHz ISM band. Moreover, BLE uses 40 channels with 2 MHz bandwidth instead of 79 tracks with 1 MHz bandwidth. As a solution, the permitted broadcast channels are reduced to three to reduce the scanning time.

```
pthread_t pid1, pid2, pid3, pid4, pid5;
t1= pthread_create(&pid[i], NULL, thread1, (void *) &input1);
if(t1){
printf("Error creating thread1%d!", i);
return -1;
t2= pthread_create(&pid[i], NULL, thread2, (void *) &input2);
if(t2){
printf("Error creating thread2%d!", i);
return -1;
t3= pthread_create(&pid[i], NULL, thread3, (void *) &input3);
if(t3){
printf("Error creating thread3%d!", i);
return -1;
t4= pthread_create(&pid[i], NULL, thread4, (void *) &input4);
if(t4){
printf("Error creating thread4%d!", i);
return -1;
t5= pthread_create(&pid[i], NULL, thread5, (void *) &input5);
if(t5){
printf("Error creating thread5%d!", i);
return -1;
```

Algorithm 2: Getting the sum value (last value of the three sensors) and publishing it to the broker to calculate the number of people in each room.

3.3.2. BLE and Positioning. Regarding the exposure to BLE devices, the services experience a new approach, and the standard itself contains multiple services. Two service profiles containing the positioning are included. Find Me Profile (FMP) and Proximity Profile (PXP). The application of FMP, which is the "Find Me Profile," is to locate the devices easier in case of loss. Two devices are needed To perform the FMP procedure, one is the locator, and the other is the target that wants to be found.

Furthermore, the server would be our target that listens for requests from the locator. Then, the target device will be notified whenever a request arrives. The notification would be a visual or auditory signal in most cases, which helps locate the device easier. The other use case of the previous notification approach is to trigger a broadcast, or through this, the device can position itself and report back with the estimated position. This standard contains another location-based service concept that is considered PXP. It defines the behavior of a device in the case of loss or connection establishment. This service is practical for a positioning platform, and it relies on the devices establishing connections. This issue leads that both PXP and FMP suffer the same problem. Where battery life is not a consideration, both profiles may be helpful in applications. iBeacon is a commercial product by Apple that provides positioning service with compatible devices and a wide range of mobile phones and computer support; however, this is not a standard profile. This is a two-way system, and its performance is the same as PXP [1].

3.3.3. Positioning Based on BLE Proximity. On a larger scale, there are many disadvantages regarding fingerprinting. Since more accuracy is unnecessary, simpler proximity or ranging systems can reduce the cost of a complete system. The

```
/* Thread-2 */
void * thread2(void * unused){
int *value;
value = (int *) unused;
if(S1 == OUT && T1 == OUT){
/*if sensorS1 and sensorT1 detect
the movement in the same
way and -1 direction it will
decrease the total value. */
room –;
}
while(1); /*loop forever*/
}
```

Algorithm 3: If sensor S1 and sensor T1 detect the movement in the same way and +1 direction, it will increase the total value.

```
/* Thread-1 */
void * thread1(void * unused){
int *value;
value = (int *) unused;
if(S1 == IN && T1 == IN){
/*if sensorS1 and sensorT1 detect the movement
in the same way and +1 direction it
will increase the total value. */
room ++;
}
while(1); /*loop forever*/
}
```

Algorithm 4: If sensor S1 and sensor T1 detect the movement in the same way and -1 direction, it will decrease the total value.

```
/* Thread-5 */
void * thread5(void * unused){
int *value;
value = (int *) unused;
if(S2 == OUT && T1 == OUT && S1 == OUT) || (S2 == IN && T1 == IN && S1 == IN){
room (–)(++);
}
while(1); /*loop forever*/
}
```

Algorithm 5: Using the thermal sensor data described in *C*.

fingerprint's implementation cost is dependent on the covered area. Therefore, a fingerprint-based system would not be the first option for indoor positioning. However, it would be an efficient idea to use the concept of fingerprinting to improve other solutions.

3.4. Beacon. The beacons frequently broadcast their I.D. and then sleep for 500 ms. The broadcast was intended only to send on the lowest of the three available channels. It is impossible to use the same program for all beacons since different hardware platforms use. However, the low complexity of the programs does not have a destructive effect on the overall system performance. The main reason for utilizing just one announcement channel is to minimize the spread created through small changes in channel properties for different frequencies.

Consequently, this reduces the fingerprint complexity. Otherwise, each beacon should be considered three times, one for each frequency. The beacons should enable to respond and provide some data to a local application where the device locate. It can be seen that this data delivery functionality and special software requirements on the devices lead to a complicated system.

3.4.1. Density and Beacon Locating Strategy. This procedure defines to receive the best performance how dense beacons are needed to locate. In a small place, the deployment of many beacons possibly only adds to noisy measurements and degrades performance. The cost and power consumption of a full-scale system should be precise before commercial deployment. To determine worst-case and best-case scenarios, this should be performed in the different radio environments to collect enough data. One way to address it is to use an overlapping quantized ranging model. The RSSI of several beacons use to determine the user's location, similar to an overlapping proximity system. However, using RSSI can be mapped to distance with low precision. This system will not impose an extra cost than a pure proximity-based one but with the same level of accuracy that practical if the positioning is only limited to the room level or similar.

3.5. MQTT Basic Definition. In IoT systems, the sensors are not autonomous to decide what to do. Sensors are responsible for sending their measurements to a central location for processing purposes in an IoT system. Then, whoever needs these data should be subscribed to them. Therefore, they would be able to process them and respond. For example, a switch can notify that it has been switched, or a temperature sensor can inform about the measured temperature. The following central location forwards all this data to the respective subscribers. This feature addresses the changes in application requirements and adds new functions.

3.6. Positioning Algorithm. This procedure performs using user position estimation through these steps. As the first step, distance estimation is performed. Two parameters are defined: RSSI (the measured power) and Tx-power (the transmitted power). A ratio between RSSI and Tx-power is defined as piecewise in Eq. (1) and Table 2 to estimate the distance. Additionally, by knowing the Beacons' positions and the distance of users from beacons using the circle's inverse equation, the position will estimate Figure 4. It can be shown by increasing the estimated distance; the received power will be decreased Figure 5.

Moreover, Figure 6 shows how people counter sensors placed in the desired building. After sensors calibration, each door will have its counting value; we use it to find the exact number of people in each dedicated place. In this simple algorithm, s1room1 = R1 is the first room in the building, and in this room, we have just one door counting because the room value is equal to the s1 (door1 calibrated sensor value). When looking at room4 = s7 − s6 − s8, which means dedicated directions of this room are inside (+) and outside is (-), and in this room, we have s7, s6, s8 (door7, door6, door8). Whenever a person comes inside room4, using s7 will increase the total number of room4; whenever a person goes outside using s6 and s8, that will decrease the total people number in room4 Figure 7. Directions are dedicated to whatever the user wants to modify the algorithm based on the dedicated paths.

$$\text{Ratio} = \frac{\text{RSSI}}{Tx - \text{power}} \quad d = \begin{cases} 10^{\text{Ratio}} & \text{Ratio} \leq 1, \\ \alpha\, \gamma \left(\beta^{\text{Ratio}}\right) & \text{Ratio} > 1. \end{cases} \tag{1}$$

3.6.1. People Counting Sensors. The main idea is to propose the people counting technique with minimum error using three sensors in each door. These three sensors should work concurrently and calibrate in a single code. The gateway can implement each sensor data using wired or wireless

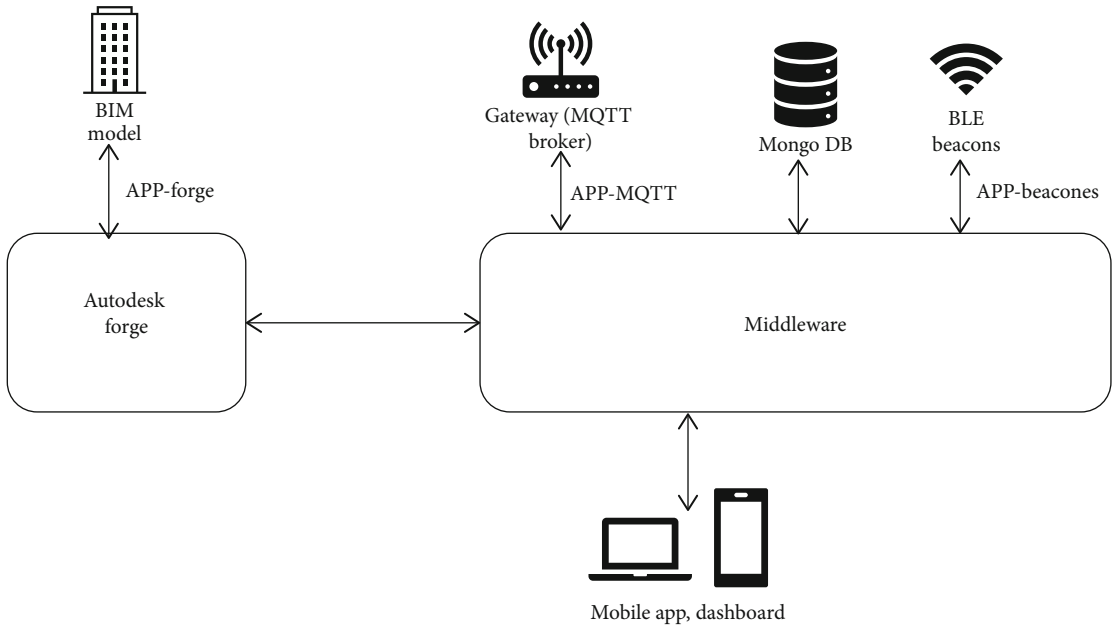


Figure 9: Image focused on the middleware section and used components and performed operations.

communication. The working principle is a simple IN-OUT principle. Whenever a person enters the room (dedicated direction for every room), the optimized data will increase the room's total number. Otherwise, it will decrease. In addition, there is a Sensor (T1), a thermal sensor (which helps not count anything else except person), and can detect the person's direction. On the left and the right side, sensors can help to detect the movement of the people. In this project, all the connected sensors to the gateway with MQTT protocol are processed with some multithread C code to compute the IN-OUT number (Figure 8).

3.6.2. Algorithm of the Counting with Three Sensors. First, subscribe to the MQTT broker to get all three-sensor data. We could subscribe and get the current data and use it. In the following, we will provide algorithms to show how to subscribe MQTT server. Here, the coding algorithm shows how the code should modify in every selected sensor in the actual testing algorithms 1–5, subscribing to the three sensors' data topic and creating five threads working concurrently in the process. We have five conditions to change the value of the final value of the three sensors.

3.7. Middleware. In this section, we focus on the middleware block and its related connection and definition. Furthermore, we have provided several images of the operation of our middleware Figure 9. Middleware block is software that serves as an interface between the end-users and other hardware components to aggregate and filter received data from the hardware components, perform information discovery, and provide access control to the end user's application.

3.7.1. Receiver. The receiver controller would be considered the crucial part of the software. The receiver works with a straightforward C-program responsible for transferring BLE packets between the gateway and the BLE host controller.

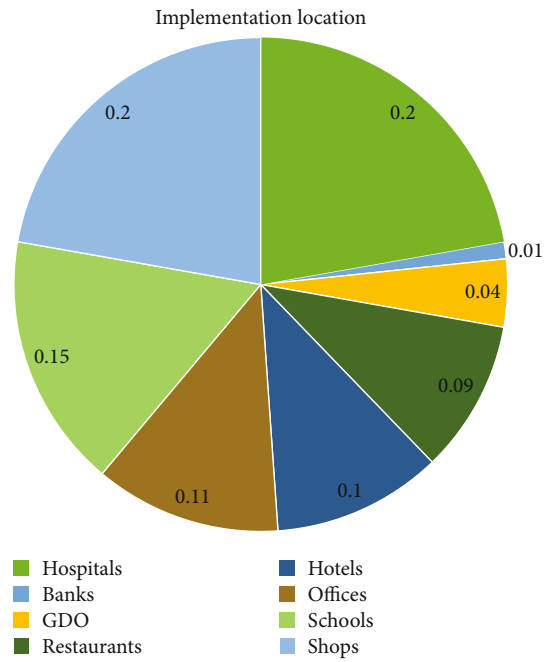


Figure 10: The pie chart illustrates information based on a survey about the most voted place to implement the project.

Packet transform performs as raw streams of data. They should decode in software to be helpful to address the essential required functions in all the experiments with the aim of code recycling maximization, for example, initialization of the host controller, initializing the radio, and scanning enabling/disabling. Support for suitable packet encoding and decoding is limited since it is time-consuming. Therefore, it does not have a significant positive point for these types of positioning systems implemented here.

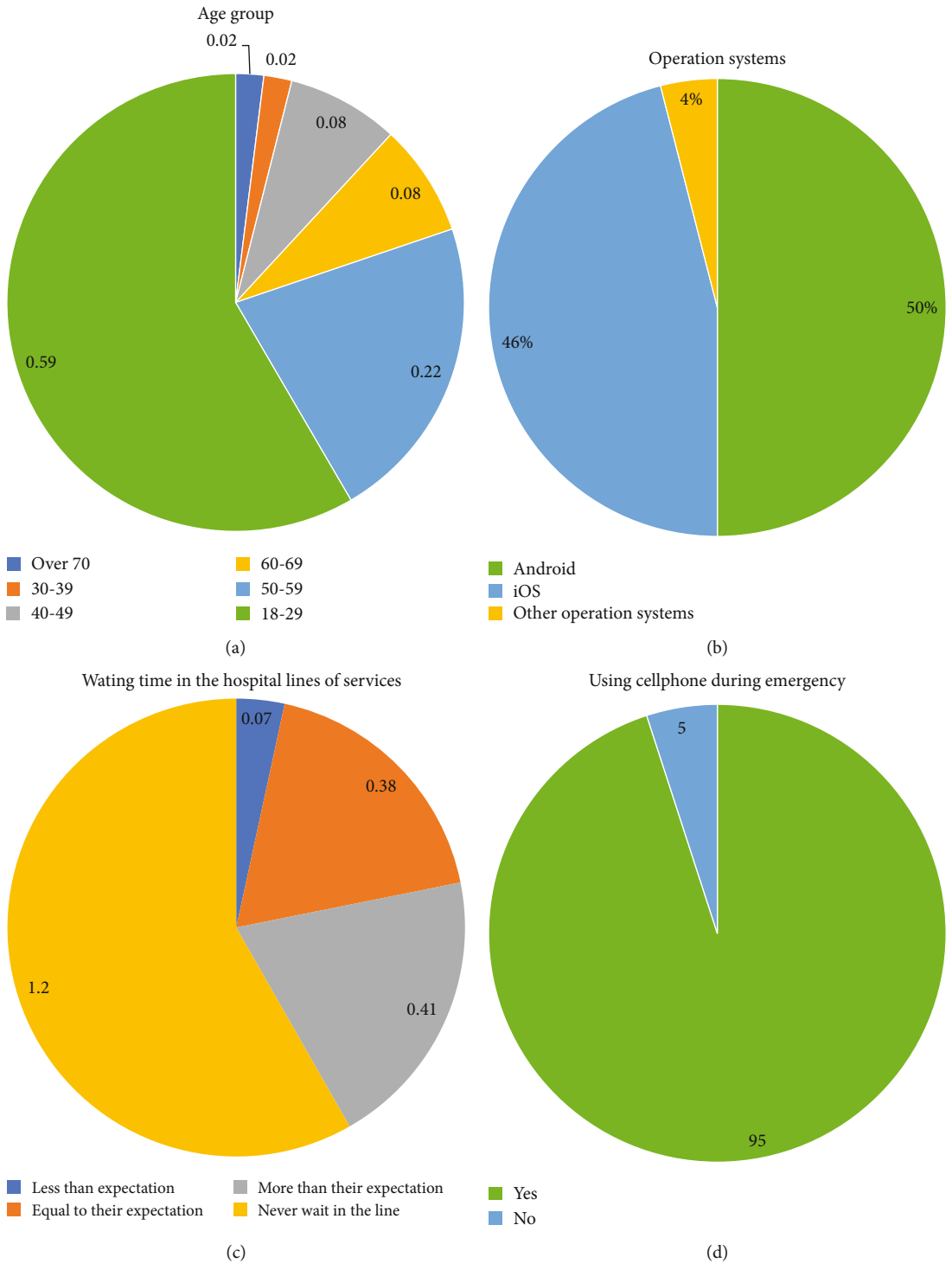


Figure 11: The pie charts a, b, c, and d depict information about preferred smartphone O.S., the age range of users, the expectation of waiting in the queues, and smartphone usage in an emergency.

4. Results and Implementation

4.1. Review Results. According to the latest surveys, we have gathered some information about our project's different aspects. Furthermore, we have found out that almost 50 percent of surveyed cases prefer to use Android, followed by 46 percent iOS. This survey has covered the data about smartphone users' age. The related pie charts illustrate that almost 60 percent of smartphone users are in the 18-29 age range, while the minor smartphone users belong to the 30–39 age groups. Regarding smartphone usage during an emergency, it can be viewed that a hefty 95 percent of cases express their

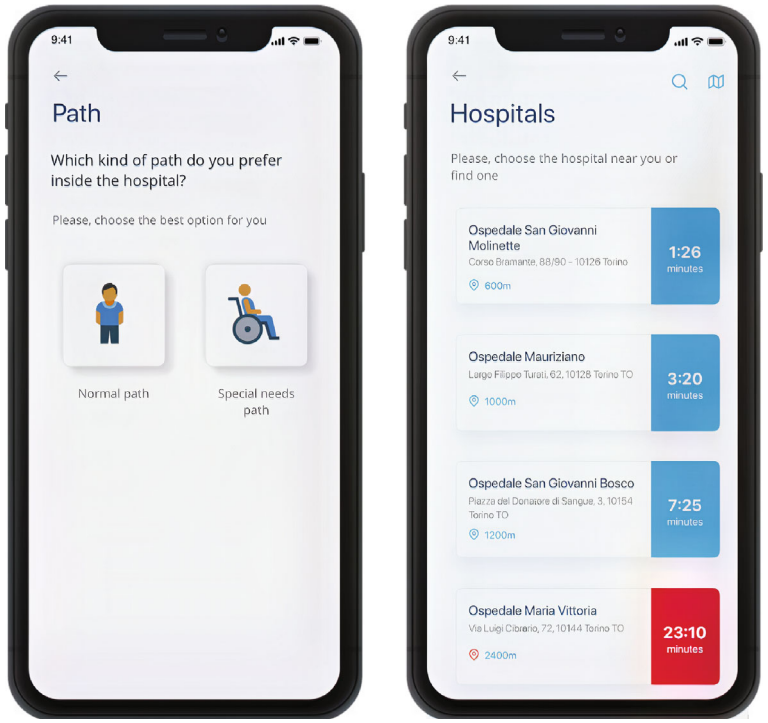


Figure 12: Interface of application.

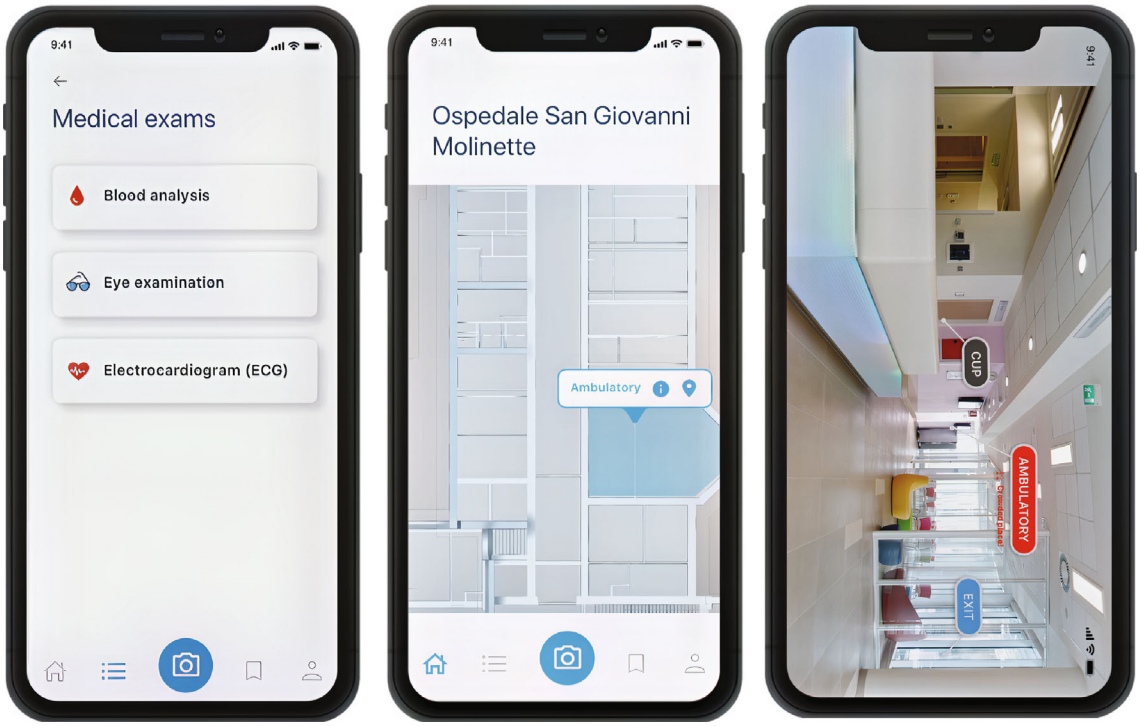


Figure 13: Interface of end-user application.

preference for using their cell phone during an emergency. In the case of waiting time that the surveyed cases must spend in hospital services lines, almost half of them have stated that is more than their expected time. Eventually, it is worth noting that the mentioned information is collected through several surveys held in Turin, Italy, Figures 10 and 11.

4.2. Platform Implementation. The proposed design was successfully implemented by integrating multiple IoT technology services and applications to create a comprehensive platform with minimal hardware usage and maximum adaptability for smart management of presented services in the hospitals. This section offers a discussion on the results

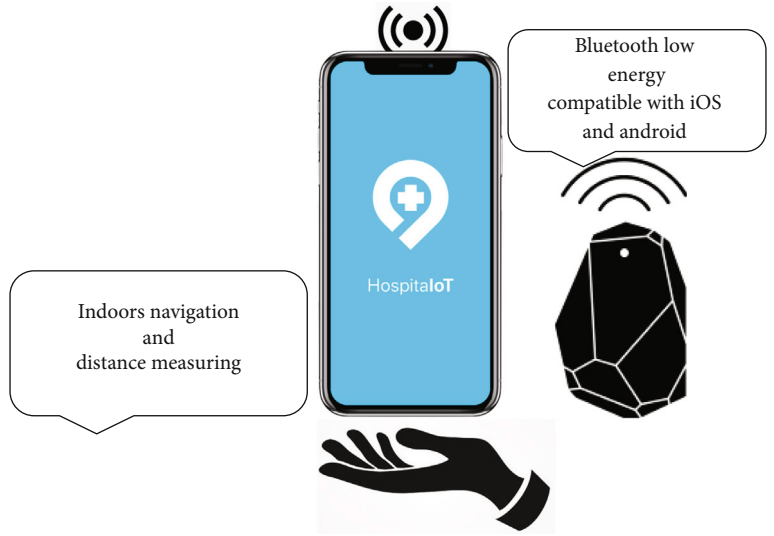


Figure 14: A general view of the project's operation when the end-user uses the application for navigation purposes.

obtained. The following objectives were satisfied with the current final application provided below. As the first result, this application provides precise navigation in the building toward the end user's desired destination through the application. This option offers a better experience for its user and decreases unnecessary commuting in the building.

As the second result of this project, the information about the point of interest can be mentioned. Through this capability of the project, end-users can access a wide range of information about the kinds of presented services and the availability of those services. The most important feature of the provided application is social distancing by notifying its users about immune distance with other people in the same area in the building. The last part of the delivered application would be highly beneficial due to the situation created by COVID-19, as Figures 12–14.

5. Conclusions

According to collected data, BLE indicates its suitability for simple positioning systems, where accuracy is not a priority. Moreover, BLE-based systems allow for low power consumption and low implementation cost. However, this is dependent on the existed local commercial offers, such as iBeacon. Although these systems are predicted to indicate the same level of functionality as proximity-based ones, ranging functionality is limited.

As the implemented system worked and provided the desired precision for a restricted set of tests, the results show that the performance of the proposed system is better than a ranging-based solution. Moreover, with some slight modifications, it can achieve a performance level like Wi-Fi-based systems with lower power consumption. Unfortunately, it is impossible to provide a detailed power consumption measurement since the available equipment cannot perform it. It is estimated that with two 3 Wh cells, the battery life for the beacons in the proposed solution can be extended to years. This study's theoretical contribution is to enhance the literature by providing a practical mobile application for the smart management of hospitals. Furthermore, the developed platform processes the buildings' information about the number of its customers and the date and time of their presence to better and smarter management of the hospital's presented services.

In addition, there is a possibility to enhance the features of the implemented system to check the space occupancy and the automatic adjustment of the environmental elements like temperature and light for better energy efficiency use.

Acronyms

GPS: Global Positioning System
MQTT: Message Queuing Telemetry Transport
TOF: Time of flight
TOA: Time of arrival
TDOA: Time difference of arrival
BLE: Bluetooth low energy
RSSI: Received signal strength indication
UWB: Ultrawideband
GSM: Global System for Mobile Communications
ISM: Industrial, Scientific, and Medical
DQPSK: Differential Quadrature Phase Shift Keying
8DPSK: 8-way differential phase shift keying
B.R.: Basic Rate
EDR: Enhanced Data Rate
GFSK: Gaussian Frequency Shift Keying

Data Availability

The experimental materials as generated software are available through the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. Larsson, *Distance Estimation and Positioning Based on Bluetooth Low Energy Technology*, 2015, http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-174857.

[2] V. G. Menon and J. Prathap, "Vehicular fog computing: challenges applications and future directions," in *Fog Computing: Breakthroughs in Research and Practice*, pp. 220–229, IGI Global, 2018.

[3] Y. Zhang, X. Deng, J. Yan, H. Su, and H. Gao, "Testing the Message Flow of Android Auto Apps," in *2019 IEEE 26th Int. Conf. on Software Analysis, Evolution and Reengineering (SANER)*, pp. 559–563, Hangzhou, China, 2019.

[4] M. M. Losavio, K. P. Chow, A. Koltay, and J. James, "The Internet of Things and the Smart City: legal challenges with digital forensics, privacy, and security," *Security and Privacy*, vol. 1, no. 3, article e23, 2018.

[5] C. Zhang, M. Kuhn, B. Merkl, M. Mahfouz, and A. E. Fathy, "Development of an uwb indoor 3d positioning radar with millimeter accuracy," in *Microwave Symposium Digest, 2006. IEEE MTT-S International*, pp. 106–109, San Francisco, CA, USA, 2006.

[6] J. Hallberg, M. Nilsson, and K. Synnes, "Positioning with bluetooth," in *10th International Conference on Telecommunications, 2003. ICT 2003*, pp. 954–958, Papeete, France, 2003.

[7] A. Kotanen, M. Hannikainen, H. Leppakoski, and T. D. Hamalainen, "Experiments on local positioning with bluetooth," in *Proceedings ITCC 2003. International Conference on Information Technology: Coding and Computing*, Las Vegas, NV, USA, 2003.

[8] A. Bekkelien, M. Deriaz, and S. Marchand-Maillet, *Bluetooth Indoor Positioning, [M.S thesis]*, University of Geneva, 2012.

[9] M. Shen, A. Liu, G. Huang, N. N. Xiong, and H. Lu, "ATTDC: an active and traceable trust data collection scheme for industrial security in smart cities," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6437–6453, 2021.

[10] I. E. A. Mansour, K. Cooper, and H. Bouchachia, "Effective live cloud migration," in *2016 IEEE 4th international conference on the future internet of things and cloud (FiCloud)*, pp. 334–339, Vienna, Austria, 2016.

[11] M. Alsanea, *The Adoption of Cloud Computing, Challenges, and Solutions, [Ph.D thesis]*, Goldsmiths, University of London, 2016.

[12] B. Wheeler, "Waggener S (2009) Above-campus services: shaping the promise of cloud computing for higher education," *Education Review*, vol. 44, pp. 10–22, 2009.

[13] G. López, L. Quesada, and L. A. Guerrero, "Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces," in *Advances in Intelligent Systems and Computing*, pp. 241–250, Springer, Champions, 2017.

[14] B. House, J. Malkin, and J. Bilmes, "The VoiceBot: a voice-controlled robot arm," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 183–192, New York, NY, USA, 2009.

[15] P. Lei, M. Chen, and J. Wang, "Speech enhancement for in-vehicle voice control systems using wavelet analysis and blind source separation," *IET Intelligent Transport Systems*, vol. 13, no. 4, pp. 693–702, 2018.

[16] Q. Ye, L. Yang, and G. Xue, "Hand-free gesture recognition for vehicle infotainment system control," in *2018 IEEE Vehicular Networking Conference (VNC)*, pp. 1-2, Taipei, Taiwan, 2018.

[17] D. K. Jain, S. Jacob, J. Alzubi, and V. Menon, "An efficient and adaptable multimedia system for converting PAL to VGA in real-time video processing," *Journal of Real-Time Image Processing*, vol. 17, pp. 11–13, 2019.

[18] C. Chakraborty, B. Gupta, S. K. Ghosh, D. K. Das, and C. Chakraborty, "Telemedicine supported chronic wound tissue prediction using classification approaches," *Journal of Medical Systems*, vol. 40, no. 3, p. 68, 2016.

[19] S. Vorapojpisut, "A lightweight framework of home automation systems based on the IFTTT model," *Journal of Software*, vol. 10, no. 12, pp. 1343–1350, 2015.

[20] X. Mi, F. Qian, Y. Zhang, and X. Wang, "An empirical characterization of IFTTT: ecosystem, usage, and performance," in *Proc 2017 Internet Measurement Conference*, pp. 398–404, New York, NY, USA, 2017.

[21] S. Rajesh, V. Paul, V. G. Menon, and M. R. Khosravi, "A secure and efficient lightweight symmetric encryption scheme for transfer of text files between embedded IoT devices," *Symmetry (Basel)*, vol. 11, no. 2, p. 293, 2019.

[22] O. Akbarzadeh, M. Khosravi, and M. Shadloo-Jahromi, "Combination of pattern classifiers based on naive Bayes and fuzzy integral method for biological signal applications," *Current Signal Transduction Therapy*, vol. 15, no. 2, pp. 136–143, 2020.

[23] O. Akbarzadeh, "Medical image magnification based on original and estimated pixel selection models," *journal of Biomedical Physics and Engineering*, vol. 10, no. 3, pp. 357–366, 2020.

[24] M. R. Khosravi, O. Akbarzadeh, S. R. Salari, S. Samadi, and H. Rostami, "An introduction to ENqI tools for synthetic aperture radar (SAR) image despeckling and quantitative comparison of denoising filters," in *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, pp. 212–221, Chennai, 2017.

[25] R. Mahmoud, T. Yousuf, F. Aloul, and I. Zualkernan, "Internet of things (IoT) security: current status, challenges and prospective measures," in *Internet Technology and Secured Transactions (ICITST), 2015 10th International Conference*, pp. 336–341, London, UK, December 2015.

[26] P. Fremantle and P. Scott, "A survey of secure middleware for the Internet of Things," *PeerJ Computer Science*, vol. 3, p. e114, 2017.

[27] T. Borgohain, U. Kumar, and S. Sanyal, "Survey of security and privacy issues of internet of things," 2015, https://arxiv.org/abs/1501.02211.

[28] O. Garcia-morchon, S. Keon, R. Hummen, and R. Struik, *Security Considerations in the IP-Based Internet of Things Draft-Garcia-Core-Security-04*, pp. 1–45, 2012, https://datatracker.ietf.org/doc/draft-garcia-core-security/04/.

[29] A. Sharma, E. S. Pilli, A. P. Mazumdar, and M. C. Govil, "A framework to manage trust in internet of things," in *Emerging trends in communication technologies (ETCT)*, pp. 1–5, Dehradun, India, November 2016.

[30] Y. Liu, S. Garg, J. Nie et al., "Deep anomaly detection for time-series data in industrial IoT: a communication-efficient on-device federated learning approach," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6348–6358, 2021.

[31] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "Blockchain for IoT security and privacy: the case study of a smart home," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 618–623, Kona, HI, USA, 2017.

[32] B. Liu, X. L. Yu, S. Chen, X. Xu, and L. Zhu, "Blockchain based data integrity service framework for IoT data," in *Proceedings -2017 IEEE 24th International Conference on Web Services, ICWS 2017*, pp. 468–475, Honolulu, HI, USA, 2017.

[33] M. Grabovica, S. Popić, D. Pezer, and V. Knežević, "Provided security measures of enabling technologies in Internet of Things (IoT): a survey," in *Zooming innovation in consumer electronics international conference (ZINC)*, pp. 28–31, Novi Sad, Serbia, June 2016.

[34] H. Javdani and H. Kashanian, "Internet of things in medical applications with a service-oriented and security approach: a survey," *Health and Technology*, vol. 8, pp. 39–50, 2018.

[35] H. Hellaoui, M. Koudil, and A. Bouabdallah, "Energy-efficient mechanisms in security of the internet of things: a survey," *Computer Networks*, vol. 127, pp. 173–189, 2017.

[36] A. Al-Gburi, A. Al-Hasnawi, and L. Lilien, *Differentiating Security from Privacy in Internet of Things: A Survey of Selected Threats and Controls Computer and Network Security Essentials*, Springer, Cham, 2018.

[37] M. A. Ferrag, L. A. Maglaras, H. Janicke, J. Jiang, and L. Shu, "Authentication protocols for Internet of Things: a comprehensive survey," *Security and Communication Networks*, vol. 2017, Article ID 6562953, 41 pages, 2017.

[38] J. Zhou, Z. Cap, X. Dong, and A. V. Vasilakos, "Security and privacy for cloud-based IoT: challenges," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 26–33, 2017.

[39] I. Khajenasiri, A. Estebsari, M. Verhelst, and G. Gielen, "A review on Internet of Things solutions for intelligent energy control in buildings for smart city applications," *Energy Procedia*, vol. 111, pp. 770–779, 2017.

[40] A. H. Alavi, P. Jiao, W. G. Buttlar, and N. Lajnef, "Internet of things-enabled smart cities: state-of-the-art and future trends," *Measurement*, vol. 129, pp. 589–606, 2018.

[41] Z. Yan, P. Zhang, and A. V. Vasilakos, "A survey on trust management for internet of things," *Journal of Network and Computer Applications*, vol. 42, pp. 120–134, 2014.

[42] F. Bao, I.-R. Chen, and J. Guo, "Scalable, adaptive and survivable trust management for community of interest based internet of things systems," in *Proc. IEEE 11th international symposium on autonomous decentralized systems (ISADS)*, pp. 1–7, Mexico City, Mexico, 2013.

[43] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: a survey, on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[44] N. S. Kim, K. Lee, and J. H. Ryu, "Study on IoT based wild vegetation community ecological monitoring system," in *Proc. 2015 7th International Conference on Ubiquitous and Future Networks*, Sapporo, Japan, July 2015.

[45] J. Y. Wang, Y. Cao, G. P. Yu, and M. Yuan, "Research on applications of IoT in domestic waste treatment and disposal," in *Proc. 11th world congress on intelligent control and automation*, Shenyang, China, 2014.

[46] J. M. Talavera, L. E. Tobón, J. A. Gómez et al., "Review of IoT applications in agro-industrial and environmental fields," *Computers and Electronics in Agriculture*, vol. 142, no. 7, pp. 283–297, 2017.

[47] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (IoT): a vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[48] Z. B. Babovic, V. Protic, and V. Milutinovic, "Web performance evaluation for internet of things applications," *IEEE Access*, vol. 4, pp. 6974–6992, 2016.

[49] Y. Wu, J. Li, J. Stankovic, K. Whitehouse, S. Son, and K. Kapitanova, "Run time assurance of application-level requirements in wireless sensor networks," in *Proc. 9th ACM/IEEE international conference on information processing in sensor networks*, pp. 197–208, Stockholm, Sweden, April 2010.

[50] S. Keshavarz, A. Abdipour, A. Mohammadi, and R. Keshavarz, "Design and implementation of low loss and compact microstrip triplexer using CSRR loaded coupled lines," *AEU - International Journal of Electronics and Communications*, vol. 111, article 152913, 2019.

[51] S. Keshavarz and N. Nozhat, "Dual-band Wilkinson power divider based on composite right/left-handed transmission lines," in *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 1–4, Chiang Mai, 2016.

[52] M. F. Ijaz, M. Attique, and Y. Son, "Data-driven cervical cancer prediction model with outlier detection and over-sampling methods," *Sensors*, vol. 20, no. 10, article 2809, 2020.

[53] M. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, "Hybrid prediction model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, synthetic minority over sampling technique (SMOTE), and random forest," *Applied Sciences*, vol. 8, no. 8, article 1325, 2018.

[54] G. Alfian, M. Syafrudin, M. Ijaz, M. Syaekhoni, N. Fitriyani, and J. Rhee, "A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing," *Sensors*, vol. 18, no. 7, article 2183, 2018.

[55] L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, "Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset," *SN computer science*, vol. 2, no. 1, p. 11, 2021.

[56] L. Garg, E. Chukwu, N. Nasser, C. Chakraborty, and G. Garg, "Anonymity preserving IoT-based COVID-19 and other infectious disease contact tracing model," *IEEE Access*, vol. 8, pp. 159402–159414, 2020.

[57] O. Akbarzadeh, M. Baradaran, and M. R. Khosravi, "IoT solutions for smart management of hospital buildings: a general review towards COVID-19, future pandemics and infectious diseases," *Current Signal Transduction Therapy*, vol. 16, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5536579, 10 pages
https://doi.org/10.1155/2021/5536579

Research Article

A Novel DBSCAN Clustering Algorithm via Edge Computing-Based Deep Neural Network Model for Targeted Poverty Alleviation Big Data

Hui Liu,[1,2] **Yang Liu**,[3] **Zhenquan Qin**,[1] **Ran Zhang**,[1] **Zheng Zhang**,[4] **and Liao Mu**[1]

[1]*School of Software Technology, Dalian University of Technology, Dalian 116024, China*
[2]*Faculty of Business and Management, Universiti Teknologi MARA Sarawak Branch Jalan Meranek, 94300 Kota Samarahan, Sarawak, Malaysia*
[3]*International School, Shenyang Jianzhu University, China*
[4]*International School of Information Science and Engineering, Dalian University of Technology, Dalian 116024, China*

Correspondence should be addressed to Yang Liu; liuyang@sjzu.edu.cn and Zhenquan Qin; qzq@dlut.edu.cn

Received 14 January 2021; Accepted 15 June 2021; Published 28 June 2021

Academic Editor: Jun Cai

Big data technology has been developed rapidly in recent years. The performance improvement mechanism of targeted poverty alleviation is studied through the big data technology to further promote the comprehensive application of big data technology in poverty alleviation and development. Using the data mining knowledge to accurately identify the poor population under the framework of big data, compared with the traditional identification method, it is obviously more accurate and persuasive, which is also helpful to find out the real causes of poverty and assist the poor residents in the future. In the current targeted poverty alleviation work, the identification of poor households and the matching of assistance measures are mainly through the visiting of village cadres and the establishment of documents. Traditional methods are time-consuming, laborious, and difficult to manage. It always omits lots of useful family information. Therefore, new technologies need to be introduced to realize intelligent identification of poverty-stricken households and reduce labor costs. In this paper, we introduce a novel DBSCAN clustering algorithm via the edge computing-based deep neural network model for targeted poverty alleviation. First, we deploy an edge computing-based deep neural network model. Then, in this constructed model, we execute data mining for the poverty-stricken family. In this paper, the DBSCAN clustering algorithm is used to excavate the poverty features of the poor households and complete the intelligent identification of the poor households. In view of the current situation of high-dimensional and large-volume poverty alleviation data, the algorithm uses the relative density difference of grid to divide the data space into regions with different densities and adopts the DBSCAN algorithm to cluster the above result, which improves the accuracy of DBSCAN. This avoids the need for DBSCAN to traverse all data when searching for density connections. Finally, the proposed method is utilized for analyzing and mining the poverty alleviation data. The average accuracy is more than 96%. The average F-measure, NMI, and PRE values exceed 90%. The results show that it provides decision support for precise matching and intelligent pairing of village cadres in poverty alleviation work.

1. Introduction

In recent years, deep learning has achieved great success in some fields; especially, the deep neural network (DNN) method has achieved good results on various tasks, such as autonomous driving, intelligent speech, and image recognition [1–4]. DNN is mainly composed of multiple convolutional layers and fully connected layers. All the layers process the input data, transmit it to the next layer, and output the results at the final layer.

The high-precision DNN model has many network layers and requires more computing resources. Currently, for intelligent application tasks, DNN is usually deployed on cloud

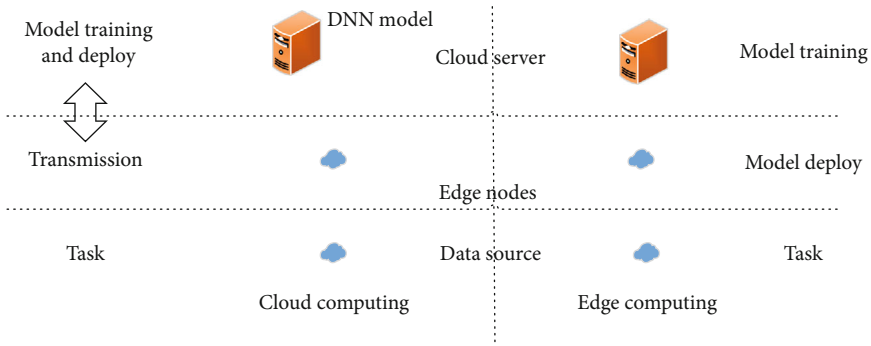


Figure 1: Comparison of cloud computing and edge computing under the deep learning model.

servers with sufficient computing resources. In this way, the data source needs to transfer the task data to the model in the cloud. The cloud computing model has the problems with high latency, low privacy, and high communication cost; so, it cannot meet the task requirements well. Some researchers attempt to deploy the cloud model to mobile devices, but due to the limited resources of mobile devices, only simple machine learning methods can be run, which results in low recognition accuracy [5–7].

As a new computing model, edge computing pushes computing power to the edge, which has attracted extensive attention from researchers [8–10]. In the edge computing scenario, the DNN model is deployed on the edge computing nodes around the device [11]. Edge computing nodes are much closer to data sources than cloud services; so, the low latency feature can be easily implemented [12]. However, the processing capacity of current edge computing devices is limited, and a single edge computing node may not be able to complete the inference task of complex network model well; so, multiple edge computing nodes are required to jointly deploy the DNN model. The main challenge of deploying DNN is how to select the appropriate computing nodes to deploy the model. Taking the segmentation of the neural network model, the computing requirements of the model and the network conditions of the edge computing nodes were taking into account; so, it optimizes the delay when multiple computing nodes jointly run the neural network model. Figure 1 shows the deep learning model under cloud computing and edge computing, respectively. In the cloud computing scenario, the data sent by the user device is uploaded to the cloud server through the network. Under the edge computing scenario, the cloud server only conducts the model training process of delay tolerance and then utilizes the trained model on the edge node. The data source transmits the task data to the edge node, and the DNN model returns the operation result for the data source.

Targeted poverty alleviation is a way to accurately identify, assist, and manage poverty alleviation targets through scientific and effective procedures in accordance with the conditions of farmers living in different poverty areas. Targeted poverty alleviation based on big data is to build a poverty portrait of poor households through the mining and analysis of poverty alleviation data and to carry out all-round identification and evaluation for poor people [13–16], so as to find out the poor population, figure out the causes of poverty, and send poverty alleviation policies. Poverty portrait generally includes poverty index, poverty characteristics, and matching assistance measures of poor households. However, the data mining method is rarely adopted for targeted poverty alleviation. Data mining analyzes data from the vast amounts of data contained in the hidden rule. The data mining includes prior knowledge-based classification, clustering without prior knowledge, association rule mining based on association rules, and intelligent algorithm based on machine learning.

Cluster analysis is an important data mining technology, which plays a very important role in data mining [17–19]. Cluster analysis technology is widely used in other research fields, such as machine learning, artificial intelligence, image processing, and cloud computing. Clustering is the process of dividing a data set into different clusters according to the similarity between data objects. The data objects belonging to the same cluster are as similar as possible. Data objects belonging to different clusters are as different as possible. So far, many different clustering algorithms have been proposed, including partition-based clustering algorithm, hierarchical clustering algorithm, density-based clustering algorithm, grid-based clustering algorithm, and fuzzy-based clustering algorithm.

The partition-based clustering algorithm uses the iterative control strategy to optimize an objective function and constantly changes the data objects in the clustering center by iterative relocation, thus improving the partitioning results in each time. The K-means algorithm is a classical partition-based clustering algorithm. Since the K-means algorithm randomly selects the clustering center, the selection of the initial clustering center will have a great impact on the clustering results, and the algorithm cannot deal with nonspherical clusters. The hierarchical clustering algorithm can handle nonspherical clusters. The Chameleon algorithm is a kind of hierarchical clustering algorithm that mixes "top down" and "bottom up" strategies. The Chameleon algorithm first constructs the k-neighborhood graph on the original data set then uses an efficient graph partition algorithm to partition the k-neighborhood graph and get the initial class cluster; finally, it merges the subclusters. However, the algorithm is sensitive to noise points and has high time complexity. The density-based clustering algorithm [20] is measured by density correlation between data points; according

to the setting threshold, the density of the density exceeds the threshold of the adjacent areas connected data cluster. Compared with partitioning-based clustering, density-based clustering can find the clustering with arbitrary shapes, and the unique outlier processing strategy can process the abnormal data effectively. The DBSCAN algorithm can discover clusters with arbitrary shape, and it is not sensitive to noise. But the time complexity is very high, and it is more sensitive to the neighborhood parameter, since different parameters can lead to different clustering results. In order to solve the problem of parameter sensitivity, Bryant et al. [21] proposed a density estimation method using the reverse nearest neighbor as the data object and used a k-neighborhood graph similar to DBSCAN for clustering.

The grid-based clustering algorithm divides the data object space into finite units according to different dimensions, and all processing takes the unit as the object [22]. In this method, the clustering operation of data sets is transformed into the blocks processing in data space, thus improving the efficiency of the algorithm. The CLIQUE algorithm [23] combines the characteristics of grid-based and density-based clustering algorithms and makes use of the priori properties of frequent patterns and association rule mining to obtain the monotony of dense elements in terms of dimension. Then, the clustering is performed by identifying dense elements. The FCM algorithm is a classic fuzzy-based clustering algorithm, which optimizes an objective function iteratively and allocates data objects according to membership matrix. Because the number of class clusters is specified in advance, if the parameter is not selected properly, it is easy to fall into local optimum.

Rodriguez et al. [24] proposed the density peak clustering (DPC) algorithm in 2014, which did not need to specify the number of class clusters in advance, and it only required fewer parameters to find nonspherical clusters and was insensitive to noise. However, the DPC algorithm also has many shortcomings including the following aspects: (1) the artificial setting of the truncation distance has a certain randomness, which has a great influence on the clustering results, and (2) when calculating the local density of data objects, it does not consider the structure difference within the data sets. When the data density difference between the clusters is large, the ideal clustering result cannot be obtained; (3) it cannot handle high dimensional data and large scale data well. In view of the shortcomings of the DPC algorithm, many algorithms have been proposed to solve the above problems. Chen et al. [25] proposed an algorithm to redefine the measurement method of truncation distance and local density by combining the concept of k-nearest neighbor, which could generate the truncation distance adaptively for any data set and make the calculation result of local density more consistent with the real distribution of data. At the same time, distance ratio contest was introduced in the decision graph to replace the original distance parameter so that the center of class cluster was more obvious in the decision graph. Guo et al. [26] adopted the idea of k-nearest neighbor to calculate the local density of data objects. Although the influence of truncation distance parameter on clustering results was largely solved, the choice of parameter k needed further study. Jin et al. [27] introduced natural neighbors to calculate the local density of the data object, thus solving the problem of parameter k selection. Li et al. [28] proposed a prominent peak clustering (PPC) algorithm based on significant density peak. The main idea of the algorithm was to divide data objects into multiple clusters and then merged the clusters with no obvious density peak to obtain accurate clustering results. Chen et al. [29] defined the local density of data objects through the law of universal gravitation and established a two-step strategy based on the first cosmic velocity to allocate the remaining data objects, so as to make the allocation of the remaining data objects more accurate. Yu et al. [30] improved the DPC algorithm by introducing weighted local density sequence and two-step allocation strategy and then used the nearest neighbor dynamic table to improve the clustering efficiency of the algorithm. Du et al. [31] introduced the k-nearest neighbor and PCA method into the DPC algorithm, which made it handle high-dimensional data well. Xu et al. [32] used the similarity index MS to allocate data points of the data set and then redefined noise points in the boundary region by the d_c-nearest neighbor method. This algorithm was suitable for data sets with high dimensional data and complex data structure, but it could not automatically determine the clustering center. Xu et al. [33] proposed a method by using grid to process large-scale data. For the above methods, they still have low efficiency when processing high dimension data.

The DBSCAN clustering algorithm uses fixed Eps and minPts (two input parameters), and the effect of processing multidensity data is not ideal. The time complexity of the algorithm is $O(N^2)$. To solve the above problems, a DBSCAN multidensity clustering algorithm based on region division is proposed. Our main contributions are as follows:

(1) First, we deploy an edge computing-based deep neural network model

(2) Then, in this constructed model, we execute data mining for poverty-stricken family

(3) In this paper, the DBSCAN clustering algorithm is used to excavate the poverty features of the poor households and complete the intelligent identification of the poor households

(4) In view of the current situation of high-dimensional and large-volume poverty alleviation data, the algorithm uses the relative density difference of grid to divide the data space into regions with different densities and adopts the DBSCAN algorithm to cluster the above result, which improves the accuracy of DBSCAN

(5) Experiments show that the algorithm can cluster multidensity data effectively and has strong adaptability to various data and better efficiency

This paper is organized as follows. In section 2, the DNN model based on edge computing for this paper is introduced.

Then, we give the detailed proposed DBSCAN algorithm in Section 3. Section 4 displays the experiment and analysis. There is a conclusion in Section 5.

2. Model Deployment Based on Edge Computing

To deploy the DNN model in an edge computing scenario, the training process of the model needs to be completed on the cloud server firstly. The structure of DNN is an ordered sequence between layers. Each layer receives the output data from the previous layer and processes it, then transmits it to the next layer. So, the DNN can be constructed into a DNN model with multiple branches, each branch is composed of a specific neural network layer, and the multiple branches constitute a complete DNN model. The edge computing nodes are composed of a variety of devices with different computing performances, which are numerous, independent, and scattered around the users. A single edge computing node can only run a simple model with low accuracy [34, 35]; so, the branching DNN model is distributed to multiple edge computing nodes. Because each branch has a different network layer structure, and this will lead to different requirements on the computing resources of the deployment nodes for each branch; that is, the deployment of the same model to different nodes will lead to different running delays. Moreover, considering the different network conditions between different edge computing nodes, the distributed deployment of the branch neural network needs to comprehensively consider the computing capacity of nodes, branch model structure, and data transmission between nodes. Therefore, it is necessary to select the best edge computing node deployment for a given DNN branching model structure.

In the edge computing scenario, given edge computing node set $E = \{e_1, e_2, \cdots, e_1, \cdots, e_m\}$ and DNN model set with n branches $D = \{d_1, d_2, \cdots, d_z, \cdots, d_n\}$, where d_z stands for the n th branch, and the n model branches require the running order, f_{ij} represent the running order of branch model i and j.

3. Proposed DBSCAN Algorithm

The DBSCAN algorithm is a classical clustering algorithm based on density. This algorithm calculates the Eps neighborhood of each data object and obtains clustering results by clustering the densified data objects into a class cluster. The DBSCAN algorithm can automatically determine the number of class clusters, find any shape of class clusters, and is not sensitive to noise data. Given a d-dimensional data set $D(i = 1, 2, \cdots, d)$, DBSCAN is defined as follows:

Definition 1. The Eps neighborhood of the data object p. The Eps neighborhood $N_{\text{Eps}}(p)$ of the data object $\forall p \in D$ is defined as the set of points contained within the region of a d -dimensional hypersphere with p as the center of sphere and Eps as the radius, i.e., $N_{\text{Eps}}(p) = \{q \in D | \text{dist}(p, q) \le \text{Eps}\}$, where D is the data set in the d-dimensional space, and dist(p, q) represents the distance between points p and q in D.

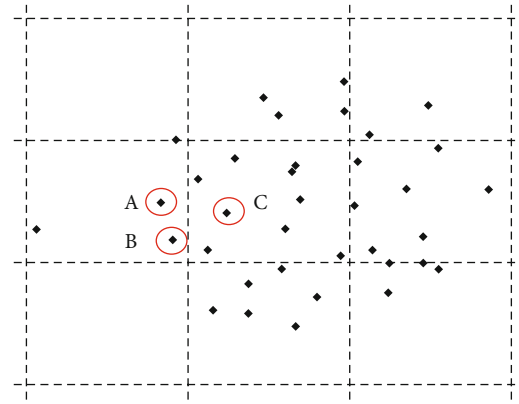


Figure 2: Boundary processing.

Definition 2. Core data objects. Given the parameters Eps and minPts, for the data object p, if the object number of p in the Eps neighborhood meets $|N_{\text{Eps}}(p)| \ge \min Pts$, then p is called the core object.

Definition 3. Directly density reachable. Given Eps and minPts, for the data object $p, q \in D$, if p satisfies the two conditions: $p \in N_{\text{Eps}}(q)$ and $|N_{\text{Eps}}(q)| \ge \min Pts$, then p is called directly density reachable about Eps and minPts. In addition, the directly density reachable does not satisfy symmetry.

Definition 4. Density reachable. Given Eps and minPts, for the data object $p, q \in D$, if there is object sequence $p_1, p_2, \cdots, p_n \in D$, $p_1 = q$, $p_n = q$, p_{i+1} is directly density reachable, then p is called density reachable about Eps and minPts. In addition, the density reachable does not satisfy symmetry.

Definition 5. Density connection. Given Eps and minPts, for the data object $p, q \in D$, if there is a data o that makes p and q are density reachable, so p and q are called density connection about Eps and minPts. Therefore, density connection satisfies symmetry.

When given Eps and minPts, the simple flow of the DBSCAN algorithm is as follows. It selects any unpartitioned data object and determines whether it is a core data object. If so, it finds all data objects with density reachable and labels these data objects as a class. Otherwise, the noise data will be judged. If it is a noise point, it will be labeled; if it is not a noise, the object will not be processed. This is repeated until all data objects have been partitioned.

Given a d-dimensional data set $D(i = 1, 2, \cdots, d)$, the number of data is N, and any dimension attribute A_i in D is bounded. Let the value of the ith dimension be in the interval $Rg_i = [l_i, h_i]$, and then $S = Rg_1 \cdot Rg_2 \cdot \cdots \cdot Rg_d$ is the d -dimensional data space. Each dimension of data space is divided into an interval with equal length and mutually disjoint to form a grid unit. These grids are left closed and right

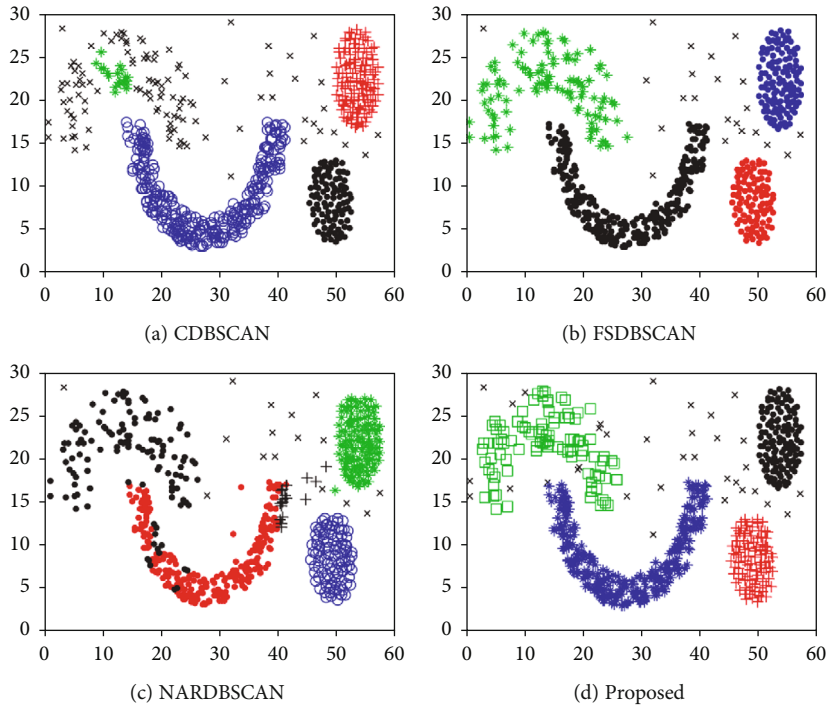


Figure 3: DS1 cluster result.

open in every dimension. In this way, the data space is divided into $\prod \text{num}_i$ super rectangular grid cells with equal volume (num_i is the number of intervals on the ith dimension in the data space). Set the grid edge length as

$$\text{length} = \alpha \cdot \sqrt[d]{\prod_{i=1}^{d} \frac{(h_i - l_i)}{N}}, \tag{1}$$

where α is the grid control factor, which is used to control the size of the grid. $a = 1.5$ is used in all experiments in this paper. According to the grid side length, the number of intervals on each dimension can be calculated as

$$\text{num}_i = \lceil (h_i - l_i)/\text{length} \rceil. \tag{2}$$

It maps each object in the data set to the corresponding grid. For each data object $X(x_1, x_2, \cdots, x_d)$, the subscript of the corresponding grid on each dimension is

$$\text{ind}_i = \lceil (x_i - l_i)/\text{length} \rceil. \tag{3}$$

For each object X in the data set, it is mapped to the corresponding grid g according to equation (3), and the number of objects in grid g is $\text{den}(g)$.

Adjacent grid is defined as if the grid g_1 and g_2 are adjacent, then $|\text{ind}_i(g_1) - \text{ind}_i(g_2)| \le (i = 1, 2, \cdots, d)$, and there are at most $3^d - 1$ adjacent grids.

The relative density difference of the grid is used; that is, two grid cells g_1 and g_2 have the density of $\text{den}(g_1)$ and $\text{den}(g_2)$, respectively, and the relative density difference of

Table 1: Comparison results with different methods.

Method	F-measure	NMI	PRE
CDBSCAN	0.8346	0.8246	0.8535
FSDBSCAN	0.9035	0.8935	0.9179
NARDBSCAN	0.9177	0.9085	0.9215
Proposed	0.9219	0.9126	0.9377

g_2 relative to g_1 is defined as

$$\text{rgdd}(g_1, g_2) = \frac{|\text{den}(g_1) - \text{den}(g_2)|}{\text{den}(g_1)}. \tag{4}$$

Formula (4) is used as a condition for grid merging. Firstly, the grid with the highest density is selected as the initial cell grid g_0, and the relative density difference $\text{rgdd}(g_0, g)$ between adjacent grid g and g_0 is calculated according to equation (4) in turn. If $\text{rgdd}(g_0, g) < \varepsilon$ (ε is a given parameter), then g and g_0 are merged. The initial unit grid g_0 becomes the merged big grid area, the initial unit grid density $\text{den}(g_0) = \text{den}(g_0) + \text{den}(g)/G_{\text{num}}$, where G_{num} represents the merged grid number; that is, the initial grid cell density is dynamic. It continues to extend outward combined grid with this method, until all the boundary grid does not meet the $\text{rgdd}(g_0, g) < \varepsilon$, because some points in the boundary grid may be boundary points for this region, as shown in Figure 2. Data points A and B are the boundary points of cluster C, which cannot be regarded as noise loss. Therefore, the boundary grid is also merged into the region (the

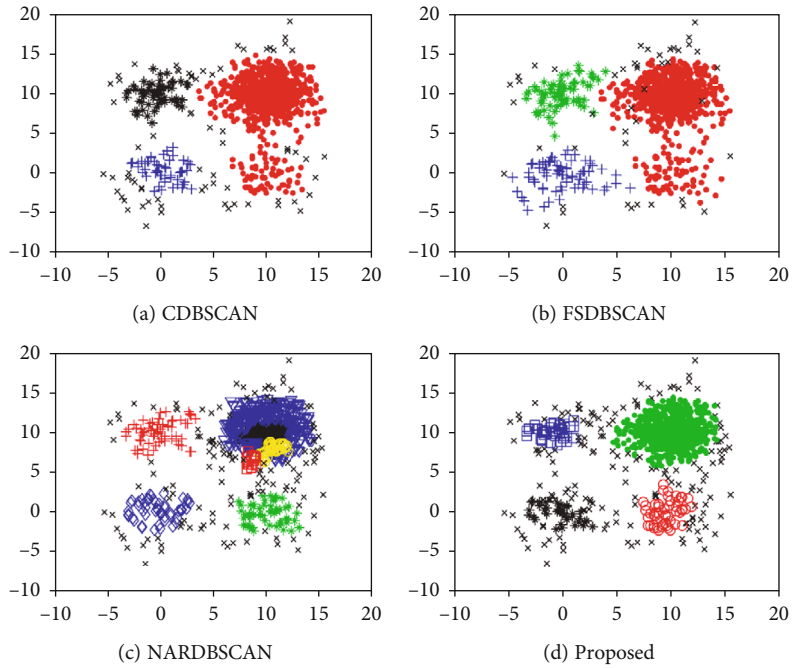


Figure 4: DS2 cluster result.

boundary grid is no longer extended outward), and the merged grid forms a region, whose data set is denoted as D_1.

Then, it takes the densest of the remaining unprocessed grids as the initial unit grid and repeats the above steps until the remaining data points can no longer be clustered. So, the data set D is divided into $D_1, D_2, \cdots, D_K$ and the initial noise point.

The above region division method has roughly divided the data set D into K data regions and noise points with different densities. Then, DBSCAN clustering is carried out for each region with different densities. According to the grid partitioning method, the grid with the largest density of the remaining grid is taken as the initial grid unit every time; so, the density from the first region to the kth region is basically decreasing.

According to the grid-based method, the grid with the largest density is used as the initial grid cell in each division. So, the density basically decreases from the first region to the Kth region. When conducting cluster, this paper inputs Eps and minPts for the first region D_1 to perform DBSCAN clustering. Eps is automatically obtained from D_2 to D_K by the following formula:

$$\text{Eps}_i = \frac{\text{Eps} \times \text{num}_1}{G_1} / \frac{\text{num}_i}{G_i}, i = 2, 3, \cdots, K, \quad (5)$$

where num_i represents the data number in D_i. G_i represents the merged grid number in the ith region.

In this paper, the data space is first roughly divided into different data regions through grid partitioning. Then, Eps is automatically obtained according to different densities for each region, and DBSCAN clustering is performed. Region partitioning reduces the unnecessary query operation of density connection in the DBSCAN algorithm and improves the efficiency. Eps of the DBSCAN algorithm is automatically obtained according to the different densities of their respective regions, which makes it more adaptable to data, especially when processing multidensity data, and the effect is better.

Table 2: Comparison results with different methods.

Method	F-measure	NMI	PRE
CDBSCAN	0.8752	0.8165	0.9069
FSDBSCAN	0.9169	0.8747	0.9344
NARDBSCAN	0.9158	0.8732	0.9317
Proposed	0.9428	0.9336	0.9572

The specific description of the proposed algorithm is as follows.

Input: dataset D, Eps, minPts, and ε.

According to equations (1)~(3), the grid is partitioned, and the data is boxed. The data set D is mapped to the partitioned grid, and the density of each grid is counted.

According to equation (4) and grid density difference parameter ε, the grids with similar densities and adjacent distances are merged. In this way, the data space is divided into different regions, and data blocks $D_1, \cdots, D_K$, and preliminary noise points are generated.

DBSCAN clustering is performed for data regions with different densities according to parameter Eps, minPts, and equation (5).

Output: clustering results.

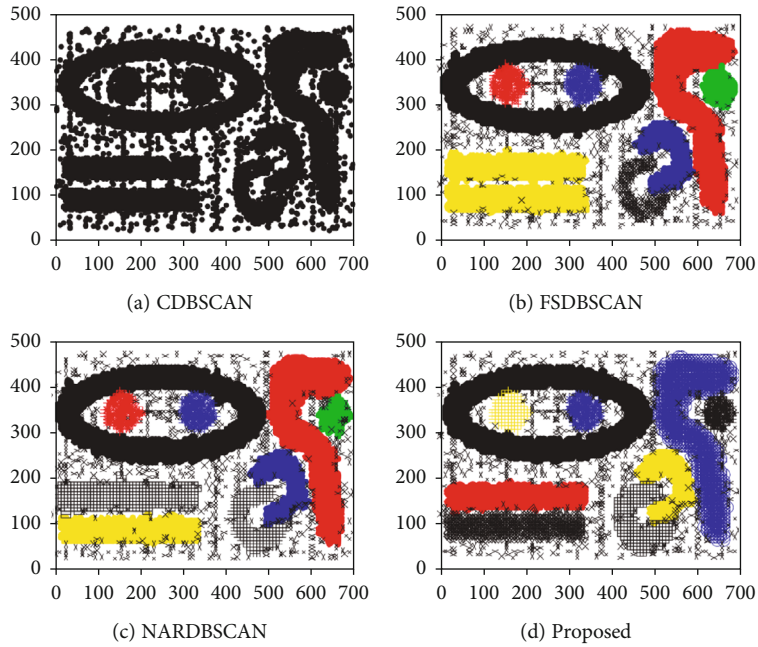


Figure 5: DS3 cluster result.

4. Experiments and Analysis

All algorithms in this paper are realized and processed by MATLAB. The experimental environment is as follows: CPU-Intel i7, Memory-4 GB, and Windows 10. We compare the proposed algorithm in this paper with the CDBSCAN [36], FSDBSCAN [37], and NARDBSCAN [38]. In this paper, three indexes including precision (PRE), normalized mutual information (NMI), and F-measure, which are widely used in clustering algorithms, are adopted as the performance measurement criteria for the clustering algorithms, where the value of NMI, F-measure, and PRE is [0,1]. The higher value denotes the better clustering result. The experiment data sets are from UCI machine learning repository (http://archive.ics.uci.edu/ml/datasets.php).

4.1. Experiment 1. Data set DS1 has four classes: noise points and multiple densities. The clustering results are shown in Figure 3.

As can be seen from Figure 3(a), data loss in low-density regions is serious when noise points are not absorbed in high-density regions. Conversely, the low-density region does not lose data, the high-density region will absorb the noise point, or the two U-shaped data on the left will be merged into a class. The clustering results in Figure 3(b) are ideal, which can identify classes with noise in the multidensity data, but the algorithm efficiency needs to be improved. In Figure 3(c), the processing effect of irregular graph boundary points is poor. In Figure 3(d), the clustering algorithm in this paper adopts the grid division method to perform DBSCAN clustering for different densities, which not only identifies four classes but also almost does not absorb noise points. In DS1 data set, the average values with different methods are shown in Table 1.

Table 1 shows the clustering results of the four algorithms on the dataset DS1. It can be seen from the comparison of F-measure, NMI, and ACC that the clustering of the proposed algorithm in this paper is higher than other algorithms on dataset DS1.

4.2. Experiment 2. This experiment uses a data set DS2 with class boundary interference. The results are shown in Figure 4.

As can be seen from Figure 4(a), CDBSCAN identifies three classes, and the class with low density and close to high density is absorbed into the high density region and absorbs more noise points. In Figure 4(b), the data center in the upper right corner has a high density, and the density gradually decreases along the edge, which is suitable for the edge data. As a result, the data in the lower right corner, which is close to the upper right corner and has a similar density, is merged into a class. In Figure 4(c), the NARDBSCAN algorithm divides the high-density region into multiple classes, and it lacks data processing from the density center to the edge. The algorithm in this paper adopts grid division and DBSCAN clustering algorithm with different parameters for different density blocks, which can well identify clusters with different density and absorb fewer noise points. Table 2 is the comparison result on data set DS2.

4.3. Experiment 3. In this experiment, data sets DS3 with large data volume, many clusters, and rich shapes are used for comparison. The data contains nine clusters. The experiment result is shown in Figure 5.

In Figure 5(b), the cluster processing of the FSDBSCAN method is ideal, but inevitably, the two classes that are close to each other are merged into one class. In Figure 5(c), low-density edges are treated as noise points, and relatively isolated points in the class are treated as noise points. In

Table 3: Comparison results with different methods on DS3.

Method	*F*-measure	NMI	PRE
CDBSCAN	0.7526	0.7359	0.7955
FSDBSCAN	0.8932	0.8132	0.8928
NARDBSCAN	0.9425	0.8872	0.9603
Proposed	0.9458	0.8893	0.9612

Figure 5(d), with the proposed method, the high-density region does not absorb too many noise points, and the low-density region is not divided or treated as noise; so, the multi-density data is well processed. Table 3 shows the detailed index values.

The above experiments mainly show the adaptability of the proposed algorithm to multidensity and arbitrary shape data clustering. The following is a quantitative analysis. The original data DS1, DS2, and DS3 are three-dimensional data, and the third dimension is the class label. If the class label in the clustering result is the same as the class label of the original data, the clustering is correct. The correct number of samples in clustering is denoted as R, and the total number of data as T [39]. The formula for calculating accuracy is

$$\text{Accuracy} = R/T \times 100\%. \tag{6}$$

The data in the following table is the average value after 20 times. DS1, DS2, and DS3 experimental results are shown in Tables 4–6.

4.4. Experiment 4. Identification of Poverty Features. This section adopts the proposed DBSCAN algorithm to process the real poverty alleviation data. The data source of poverty alleviation comes from a prefecture-level city, including 11,423,500 rural population, 196,700 rural households, and 68,000 poverty-stricken households. The original data includes the data of archived card, the data of visit, the data of agricultural cloud project, the data of education, and health and sanitation departments. According to the designed poverty alleviation data indicator system, the data of poverty alleviation can be filled with ETL tool. Finally, the proposed DBSCAN algorithm in this paper is applied to analyze the poverty alleviation index data to intelligently identify the poverty characteristics of the poor households and carry out the corresponding visual display and analysis to provide decision support for the regional assistance work.

In this paper, the proposed DBSCAN clustering algorithm based on local sensitive hash is applied to complete the mining and analysis of the preprocessed poverty alleviation index data matrix to identify the poverty characteristics of the poor households in different regions. Feature classification is carried out by clustering algorithm, and 8 data clusters are finally identified by consensus with each cluster corresponding to a feature category. For each data cluster, it takes the central data object as the cluster. The index whose value exceeds the threshold value of 0.7 is the impoverishment index of the data object. The poverty characteristics identified in this paper are compared and corresponded to the causes of poverty in the registered cards of poor households. On the one hand, it refines and complements the general causes of poverty in the archival registration cards to facilitate the matching of assistance measures and assistance cadres in the later period. On the other hand, the identification results of this paper and the causes of poverty can be used to verify each other.

Table 4: DS1 result.

Method	Time/ms	Accuracy/%
CDBSCAN	262.9	82.63
FSDBSCAN	218.5	96.3
NARDBSCAN	114.5	95.28
Proposed	45.9	99.24

Table 5: DS2 result.

Method	Time/ms	Accuracy/%
CDBSCAN	208.4	85.84
FSDBSCAN	125.8	93.64
NARDBSCAN	63.1	94.82
Proposed	38.7	96.87

Table 6: DS3 result.

Method	Time/ms	Accuracy/%
CDBSCAN	669.4	88.95
FSDBSCAN	263.7	94.62
NARDBSCAN	90.7	96.27
Proposed	55.5	99.36

Table 7: Poverty feature recognition accuracy.

Method	Accuracy	Time
CDBSCAN	91.0%	One week
FSDBSCAN	92.3%	4 days
NARDBSCAN	94.5%	2.5 days
Proposed	96.5%	1358s

Table 7 shows the comparison accuracy between the poverty characteristics identified in this paper and the causes of poverty caused by registered card. We give the overall comparison results. Because this article identifies poverty characteristics compared with cross tent card of poverty causes more refined, index, and rich content, the farmers that have the same poverty characteristics may correspond to multiple poverty reasons. When calculating the accuracy, for the poverty characteristics of a certain household, some or all of the poverty characteristics correspond to the causes of poverty caused by registered card, and we believe that the recognition result is consistent with the result of registered card.

5. Conclusions

Targeted poverty alleviation aims to build a well-off society in an all-round way. However, the current poverty alleviation work mainly relies on the establishment of documents, visiting by village cadres, and so on, which costs a lot of manpower and material resources, and lacks scientific and effective means with supervision management. This paper mainly studies the intelligent identification method of poverty characteristics for poor households, designs an intelligent identification scheme of poverty characteristics, and further provides data reference and guidance for accurate matching of assistance measures and pairing of assistance cadres. However, if the dimension of the datasets is very higher, the cluster effect is not perfect with the current experimental environment. In the future, we will adopt newly deep learning methods to improve the cluster effectiveness. The improved and advanced clustering technologies will be applied in targeted poverty alleviation of the poverty counties in China.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] Q. Zhang, C. Bai, Z. Chen et al., "Deep learning models for diagnosing spleen and stomach diseases in smart Chinese medicine with cloud computing," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 7, p. 1, 2021.

[2] P. Li, Z. Chen, L. Yang, Q. Zhang, and M. J. Deen, "Deep convolutional computation model for feature learning on big data in internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 790–798, 2018.

[3] L. Zhao, Z. Chen, Y. Yang, V. C. M. Leung, and Z. Jane Wang, "Incomplete multi-view clustering via deep semantic mapping," *Neurocomputing*, vol. 275, pp. 1053–1062, 2018.

[4] J. Gao, P. Li, and Z. Chen, "A canonical polyadic deep convolutional computation model for big data feature learning in Internet of Things," *Future Generation Computer Systems*, vol. 99, pp. 508–516, 2019.

[5] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, 2019.

[6] R. Travadi and S. Narayanan, "Total variability layer in deep neural network embeddings for speaker verification," *IEEE Signal Processing Letters*, vol. 26, no. 6, pp. 893–897, 2019.

[7] P. Li, Z. Chen, L. Yang, J. Gao, Q. Zhang, and M. J. Deen, "An incremental deep convolutional computation model for feature learning on industrial big data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1341–1349, 2019.

[8] Z. Ning, X. Kong, F. Xia, X. Wang, and W. Hou, "Green and sustainable cloud of things: enabling collaborative edge computing," *IEEE Communications Magazine*, vol. 57, no. 1, pp. 72–78, 2019.

[9] A. A. Laghari, H. He, A. Khan, N. Kumar, and R. Kharel, "Quality of experience framework for cloud computing (QoC)," *IEEE Access*, vol. 6, pp. 64876–64890, 2018.

[10] L. Liu, G. Han, Z. Xu, L. Shu, B. Peng, and M. Martinez-Garcia, "Predictive boundary tracking based on motion behavior learning for continuous objects in industrial wireless sensor networks," *IEEE Transactions on Mobile Computing*, 2021.

[11] A. Laghari, H. He, S. Karim, H. A. Shah, and N. K. Karn, "Quality of experience assessment of video quality in social clouds," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 8313942, 10 pages, 2017.

[12] G. Han, H. Wang, H. Guan, and M. Guizani, "A mobile charging algorithm based on multicharger cooperation in internet of things," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 684–694, 2021.

[13] Y. Zhou, Y. Guo, Y. Liu, Y. Li, and W. Wu, "Targeted poverty alleviation and land policy innovation: some practice and policy implications from China," *Land Use Policy*, vol. 74, pp. 53–65, 2018.

[14] Q. Zhang and T. Zhang, "Advances and practices for targeted poverty alleviation in China," *China Economic Transition*, vol. 1, no. 1, pp. 134–141, 2018.

[15] W. Luo, Y. Tang, and X. Zhang, "Analysis on factors influencing farmers' participation behavior in rural tourism targeted poverty alleviation:based on embedding social structure theory," *Journal of Hunan Agricultural University*, vol. 20, no. 5, pp. 24–30, 2019.

[16] C. Chen and J. Pan, "The effect of the health poverty alleviation project on financial risk protection for rural residents: evidence from Chishui City, China," *International Journal for Equity in Health*, vol. 18, no. 1, p. 79, 2019.

[17] L. Teng, H. Li, and S. Yin, "Modified pyramid dual tree direction filter-based image denoising via curvature scale and nonlocal mean multigrade remnant filter," *International Journal of Communication Systems*, vol. 31, no. 16, article e3486, 2018.

[18] M. Nagano, A. Komesu, and H. Miyata, "An evolutionary clustering search for the total tardiness blocking flow shop problem," *Journal of Intelligent Manufacturing*, vol. 30, no. 4, pp. 1843–1857, 2019.

[19] M. Parmar, D. Wang, X. Zhang et al., "REDPC: A residual error-based density peak clustering algorithm," *Neurocomputing*, vol. 348, pp. 82–96, 2019.

[20] X. Cui, J. Wang, F. Wu et al., "Extracting main center pattern from road networks using density-based clustering with fuzzy neighborhood," *International Journal of Geo Information*, vol. 8, no. 5, p. 238, 2019.

[21] A. Bryant and K. Cios, "RNN-DBSCAN: a density-based clustering algorithm using reverse nearest neighbor density estimates," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 6, pp. 1109–1121, 2018.

[22] L. Wang, S. Ding, Y. Wang, and L. Ding, "A robust spectral clustering algorithm based on grid-partition and decision-graph," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 5, pp. 1243–1254, 2021.

[23] M. Hassani, Y. Kim, and T. Seidl, "Subspace MOA: subspace stream clustering evaluation using the MOA framework," in *International Conference on Database Systems for Advanced Applications*, Wuhan, China, 2013.

[24] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[25] J. Chen and P. Yu, "A domain adaptive density clustering algorithm for data with varying density distribution," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2310–2321, 2021.

[26] Z. Guo, T. Huang, Z. Cai, and W. Zhu, "A new local density for density peak clustering," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Melbourne, VIC, Australia, 2018.

[27] H. Jin and X. Z. Qian, "Optimized density peak clustering algorithm by natural nearest neighbor," *Journal of Frontiers of Computer Science and Technology*, vol. 13, no. 4, pp. 711–720, 2019.

[28] L. Ni, W. Luo, W. Zhu, and W. Liu, "Clustering by finding prominent peaks in density space," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 727–739, 2019.

[29] H. Chen, M. Ge, and Y. Xue, "Clustering algorithm of density difference optimized by mixed teaching and learning," *SN Computer Science*, vol. 1, no. 3, 2020.

[30] D. Yu, G. Liu, M. Guo, X. Liu, and S. Yao, "Density peaks clustering based on weighted local density sequence and nearest neighbor assignment," *IEEE Access*, vol. 7, pp. 34301–34317, 2019.

[31] M. Du, S. Ding, and H. Jia, "Study on density peaks clustering based on k-nearest neighbors and principal component analysis," *Knowledge-Based Systems*, vol. 99, pp. 135–145, 2016.

[32] X. Xu, S. Ding, H. Xu, H. Liao, and Y. Xue, "A feasible density peaks clustering algorithm with a merging strategy," *Soft Computing*, vol. 23, no. 13, pp. 5171–5183, 2019.

[33] X. Xu, S. Ding, and Z. Shi, "An improved density peaks clustering algorithm with fast finding cluster centers," *Knowledge-Based Systems*, vol. 158, pp. 65–74, 2018.

[34] Z. Liao, G. Han, H. Wang, and L. Liu, "Multistation-based collaborative charging strategy for high-density low-power sensing nodes in industrial Internet of Things," *IEEE Internet of Things Journal*, vol. 8, no. 9, pp. 7575–7588, 2021.

[35] A. Laghari, A. K. Jumani, and R. A. Laghari, "Review and state of art of fog computing," *Archives of Computational Methods in Engineering*, pp. 1–13, 2021.

[36] T. Tran, K. Drab, and M. Daszykowski, "Revised DBSCAN algorithm to cluster data with dense adjacent clusters," *Chemometrics and Intelligent Laboratory Systems*, vol. 120, pp. 92–96, 2013.

[37] D. Abdolzadegan, M. Moattar, and M. Ghoshuni, "A robust method for early diagnosis of autism spectrum disorder from EEG signals based on feature selection and DBSCAN method," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 482–493, 2020.

[38] Z. Han, M. Cheng, F. Chen, Z. Deng, and Y. Wang, "A spatial load forecasting method based on DBSCAN clustering and NAR neural network," *Journal of Physics Conference Series*, vol. 1449, article 012032, 2020.

[39] Z. Xu, G. Han, H. Zhu, L. Liu, and M. Guizani, "Adaptive DE algorithm for novel energy control framework based on edge computing in IIoT applications," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 5118–5127, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5581259, 14 pages
https://doi.org/10.1155/2021/5581259

Research Article

Evaluation of Congestion Aware Social Metrics for Centrality-Based Routing

Muhammad Arshad Islam,[1] **Muhammad Azhar Iqbal**,[2] **Muhammad Aleem**,[1] **Zahid Halim**,[3] **Gautam Srivastava**,[4,5] **and Jerry Chun-Wei Lin**[6]

[1]*National University of Computer and Emerging Sciences-FAST, Islamabad, Pakistan*
[2]*School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China*
[3]*Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Swabi, Pakistan*
[4]*Department of Math and Computer Science, Brandon University, Canada*
[5]*Research Centre for Interneural Computing, China Medical University, Taiwan*
[6]*Western Norway University of Applied Sciences, Norway*

Correspondence should be addressed to Jerry Chun-Wei Lin; jerrylin@ieee.org

Received 13 February 2021; Accepted 4 June 2021; Published 21 June 2021

Academic Editor: Nawab Muhammad Faseeh Qureshi

Opportunistic networks utilize pocket switching for routing where each node forwards its messages to a suitable next node. The selection of the forwarder node is crucial for the efficient performance of a routing protocol. In any opportunistic network, some nodes have a paramount role in the routing process and these nodes could be identified with the assistance of the existing centrality measures available in network theory. However, the central nodes tend to suffer from congestion because a large number of nodes that are relatively less central attempt to forward their payload to the central nodes to increase the probability of the message delivery. This paper evaluates mechanisms to transform the social encounters into congestion aware metrics so that high-ranking central nodes are downgraded when they encounter congestion. The network transformations are aimed at aggregating the connectivity patterns of the nodes to implicitly accumulate the network information to be utilized by centrality measures for routing purposes. We have analyzed the performance of the metrics' computed centrality measures using routing simulation on three real-world network traces. The results revealed that betweenness centrality along with the congestion aware network metrics holds the potential to deliver a competitive number of messages. Additionally, the proposed congestion aware metrics significantly balance the routing load among the central nodes.

1. Introduction

Communication in opportunistic networks is dependent on the mobility patterns of the wireless nodes that constitute the network. An opportunity to exchange data among the network is created whenever two nodes come across the communication range of each other. The wireless nodes exhibit a store-carry-forward mechanism wherein they hold the characteristics of both data-mule and router [1]. This intermittent nature of the opportunistic network causes the delivery time of a message to vary from a few seconds to several days depending on the node characteristics (mobility model), network characteristics (network density and network diameter), and message characteristics (message size and the distance between source and destination) [2].

An efficient opportunistic network routing protocol accurately predicts connectivity pattern using available node-node link characteristics such as contact count, contact duration intercontact time (time elapsed between two consecutive contacts) [1], frequency, and duration of the contacts that are created as a result of the social interaction of the users of the nodes. Due to this versatile nature of opportunistic networks, extracting an accurate routing metric for different kinds of opportunistic networks becomes a strenuous process. However, in any network, a few nodes are considered central nodes (also known as Hub) since they are

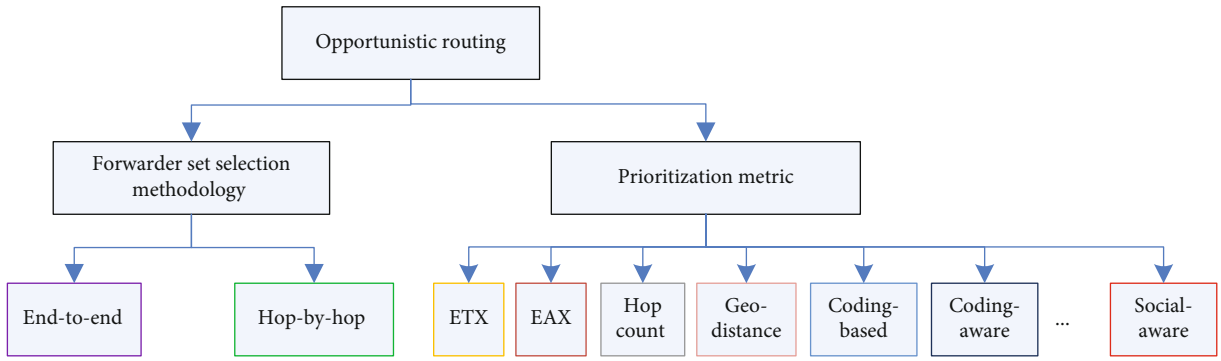


Figure 1: Components of opportunistic routing.

playing a pivotal role in routing, e.g., these central nodes contribute more in the process of message forwarding than the other nodes [3].

The fundamental notion of opportunistic routing consists of two main components, i.e., forwarding methodology and prioritization as shown in Figure 1. Forwarding methodology can be based on hop-by-hop selection of forwarder/-central nodes or end-to-end selection of forwarder node set. On the other hand, the prioritization is based on a metric (i.e., node sociality, Expected Transmission Count (ETX), geo-distance, hop count, etc.) dependent on the particular nature of a wireless network or wireless network application and it ultimately suppresses undue packet forwarding.

Network centrality is considered a vital tool for network analysis to identify because such central nodes can play a key role in disseminating information in the network. Most of the centrality measures are defined for static networks (contact information among nodes does not change with time). This work focuses on evaluating the routing performance of the network metrics that are adapted to implicitly harness the congestion-related information of the links between mobile nodes using centrality measures. The congestion aware nature of the investigated metrics manipulates the centrality ranking in such a way that a high-rank mobile node will be ranked lower than its counterparts if it carries higher traffic volume.

The remainder of the paper is organized as follows: Section 2 presents existing state-of-the-art works that use centrality measures for routing. Section 3 presents the metrics that are used to transform the opportunistic networks. Details of the network traces along with the simulation setup parameters are discussed in Section 4. Section 5 presents the results and discusses the performance of simulated network metrics. Conclusion and future work are presented in Section 6.

2. Related Work

The process of central node identification has been utilized in multiple works to make routing decisions in different kinds of communication networks. Literatures acknowledge the synergy among social network tools like centrality measures and ad hoc networks as a fertile research area that has the potential for designing network routing protocols for opportunistic networks [4]. Researchers have the choice among various centrality measures available in network theory [5]. The centrality of a node in a network can be determined by using various centrality measures such as degree centrality [6], betweenness centrality [5], eigenvector centrality [7], page-rank [8], hub centrality [9], and centrality [10]. The computation process of most of the centrality measures is centralized and required complete network information. However, in the context of opportunistic networks, it is not practical for a node to have the complete information of the network and to calculate various centrality measures ranking a node concerning its suitability for routing [10]. Moreover, due to the dynamic nature of opportunistic networks, the contemporary centrality measure cannot be helpful to calculate node centrality [11].

The existing centrality measures have been applied to address multiple issues of ad hoc network such as routing [12], congestion [13, 14], and energy conservation [1]. Wang et al. proposed a scheme for superior forwarding node selection that is based on the concept of value strength that relies on social network structure [15].

The authors simulated the protocol on real-world traces extracted from Flickr to show high information converge ratio. Zhu et al. [16] have presented a routing protocol "ZOOM" for opportunistic forwarding in vehicular networks that uses network centrality metrics in the absence of primary routing information, i.e., intercontact time. The concept of centrality-based community is employed for opportunistic routing where a message is pushed towards a central community with the understanding that nodes in such a community have a high probability to get connected with the destination of the message [17]. It has also been argued that nodes with a high value of betweenness centrality are susceptible to face traffic congestion as well as energy depletion because of their high probability of participation in the routing process [18]. Miralda et al. [19] have employed a variant of betweenness centrality based on fuzzy logic with the aim of energy conservation that is crucial in Io nodes in opportunistic networks. They argue that nodes with high local betweenness centrality are prone to consume more energy and face buffer occupancy problems during the routing process, and consequently, distributing the routing process with low local betweenness centrality can help in conserving the energy.

The effectiveness of closeness and betweenness centrality to identify the influential nodes for routing in opportunistic networks is demonstrated in [20]. Sivalingam and Chellappan [21] have used the concept of entropy for routing purposes in tactical wireless networks. Entropy is defined in terms of the downstream degree of a vertex, which is the number of eligible vertices for forwarding. The investigation also includes the correlation among centralities as a function of network connectivity and network mobility showing that the closeness centrality (relevant to the shortest path) has obtained a higher correlation with degree-based centrality measure as compared to betweenness centrality. The research community agrees that network-theoretic concepts can be useful for the identification of influential nodes for routing in infrastructure-less environments. However, this brings about the challenge of how centrality measures for static networks can be transformed for making accurate routing decisions in opportunistic networks using metrics that preserve the link characteristics between the nodes for congestion handling. The list of link characteristics includes node connectivity count, link error rate, link duration, and hidden node problems. The focus of this work is to investigate the congestion aware metrics using node contact patterns to enhance the routing performance in the opportunistic network.

Gap analysis: we have discussed several recent studies that have focused on improving routing performance and reducing energy consumption using centrality metrics. The focus of these studies is mostly computing centrality measures with minimum network overhead that requires global network knowledge. However, none of the recent works have investigated the implicit congestion avoidance capabilities of the network metrics that can be sensed and shared among the network devices with little or no overhead. This paper investigates three such network metrics by integrating them with multiple centrality measures to gauge their congestion awareness.

3. Opportunistic Network Metrics

The metrics are aimed at facilitating the computation of centrality measures while maintaining the temporal characteristics of the contact among the network nodes. The aforementioned centrality measures have been described concerning the static networks. However, opportunistic networks are dynamic where nodes may join and leave the network. We can transform the dynamic link behavior of the network nodes using the following metrics to analyze their performance for opportunistic network routing. We have divided the presented metrics into two classes, i.e., congestion oblivious and congestion aware.

3.1. Congestion Oblivious Metric. Most of the existing literature relies on the congestion oblivious metrics that are not affected by the current traffic load of the routing nodes.

Aggregate network: it is the simplest metric for the centrality computation for dynamic networks. An opportunistic network can be seen as a sequence of static graphs at a particular point in time as shown in Figure 2(a). A simple network is created aggregating all edges that existed at any point in time [22] as shown in Figure 2(b). An aggregated graph for a dynamic network can be represented as an $n \times n$ adjacency matrix A, where each element a_{ij} is defined as follows:

$$\alpha_{ij} = \begin{cases} 1 & \text{nodes } i\&j \text{ made a contact,} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

The primary advantage of this method is simplicity. However, aggregating all edges may result in losing temporal information that is vital for routing decisions in opportunistic networks. A connection between any two nodes that once came in contact with each other will be represented as a permanent connection.

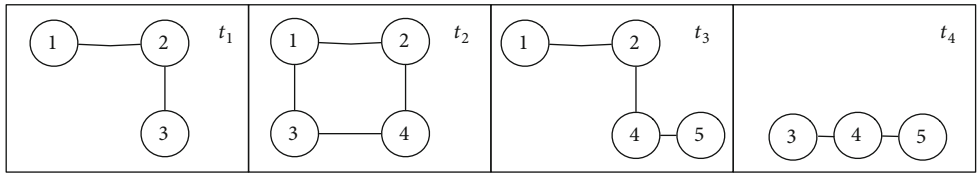
Contact count: it represents a weighted network with edges representing the number of contacts that occurred between two nodes. This metric will relate two frequently contacting (two nodes contact each other when they are in their wireless range) nodes stronger than those that have contacted each other less frequently. Contact count graph can be represented as a weighted adjacency matrix ContCnt where each element η_{ij} is defined as the total contact count between nodes i and j. This metric allocates weight to contacts that occur frequently; however, it does not favor contacts occurring at regular intervals. Two nodes that connect infrequently but on regular basis (daily) over a longer span will be given less priority as compared to the nodes that get connected very frequently over a shorter period.

3.2. Congestion Aware Metrics. Duration-based metrics have not been investigated to handle congestion. The focus of this work is to evaluate the congestion aware metrics that affected implicitly the traffic load of a node, and they can be used to lower the routing suitability of the nodes that are facing higher routing load [23].

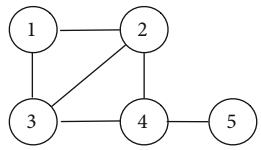
Average contact duration: it represents a weighted network with edges representing the average duration of the contacts between any two nodes. The longer the average contact duration between two nodes, the stronger is the relationship between them and vice versa. The contract duration of two nodes will be reduced if they participate in large volume transmissions. Thus, their centrality rank will be lowered due to congestion. The average duration network can be represented as a weighted adjacency matrix Dur where each element λ_{ij} is defined as follows:

$$\lambda_{ij} = \frac{\sum_{k=1}^{\eta_{ij}} \text{contactduration}_k}{\eta_{ij}}. \quad (2)$$

Intercontact time: it represents a weighted network with edges representing the mean time elapsed between two consecutive contacts of two particular nodes. The smaller the intercontact time between two nodes, the stronger is the relationship between them and vice versa. This is a duration-based metric that will affect the relationship between two nodes if they encounter network congestion. The duration of the contact time of the nodes will be reduced, and



(a) Snapshots of an opportunistic network with 5 nodes at time t_1, t_2, t_3, t_4



(b) Aggregated network from interval t_1 to t_4

Figure 2: Opportunistic network representation as an aggregated network.

correspondingly, intercontact time will increase. Intercontact time network of nodes i and j with η_{ij} contact count can be represented as weighted adjacency matrix ICDur where each element μ_{ij} is defined as follows:

$$\mu_{ij} = \frac{\sum_{k=1}^{\eta_{ij}} \text{intercontactDur}_k}{\eta_{ij}}. \quad (3)$$

The metric considers the mean of all the intercontact time duration elapsed among all contacts of two nodes. The behavior of this metric is somewhat similar to the contact count. This metric will downgrade the relationship between two nodes that either stop contacting each other or start to incur longer delays.

All of the metrics except aggregate graph are dynamic in nature, and two metrics, i.e., average contact duration and intercontact time are sensitive to congestion faced by the involved nodes. Considering dynamic metric, the node ranking is not static and the ranking of a central node bounds to degrade with passing time when the node encounters congestion while congestion sensitive metrics are employed.

4. Experimental Setup

We have considered three different kinds of datasets, namely, MIT cell tower, MIT Bluetooth, and IBM access points, and all of these have been obtained from Community Resource for Archiving Wireless Data at Dartmouth (http://www.crawdad.org/). The selected dataset promises to represent the real-life mobility pattern of users while grabbing the basic social contact behavior. The motivation behind choosing these three traces is to analyze the range of spectrum between dense and sparse networks. Two of the data traces have been synthesized from reality mining project [24] from MIT spans 19 months, i.e., February 2004 to August 2005, whereas the third data trace consists of the SNMP logs for one month from an IBM campus [25]. Since the contact duration of MIT reality mining is longer than IBM trace, we have filtered the MIT data to match the time span of IBM traces.

The sparse network is extracted from Bluetooth logs of MIT traces (MITBT) where each node scans for active Bluetooth neighbors in the interval of every five minutes and stores the duration of contact times. For the sake of comparison with other traces and simplicity, we have limited our experiments to one month of connectivity trace, where any visible Bluetooth device was considered a candidate connection. Reduction of the trace time span has been considered on the basis of connectivity times, i.e., one month, where nodes have maximum connectivity in terms of time duration. The highest connectivity period, i.e., November 2004, showed 1858 Bluetooth nodes suggesting a huge number of undesignated nodes as compared to the designated 81 nodes that were designated to gather the data. It is noteworthy that a few undesignated devices had more connectivity and interaction with the network than the designated nodes.

In the case of IBM Access Point trace, Simple Network Management Protocol (SNMP) is used to poll access points (AP) every 5 minutes, from July 20, 2002, through August 17, 2002 [25]. The total of 1366 devices has been polled over 172 different access points during approximately 4 weeks. We have extracted the traces of 928 devices after discovering the existence of 3 clusters in this network and then choosing the biggest cluster with respect to node count. The biggest cluster has been identified by analyzing the connectivity pattern among devices. The 3 extracted clusters represent the devices belonging to 3 buildings, and the biggest cluster is considered for the simulations. Since the authors of the dataset [25] have polled the access point for connected devices every 5 minutes, we assume that the snapshot data remains constant for the next 5 minutes to turn these samples into continuous data. In the rare cases where this would cause an overlap with another snapshot from another access point, we assume that the transition happens halfway between the two snapshots. It is also assumed that two nodes that are connected to one access point during the overlapping time period are connected to each other. Thus, key features of such network are low mobility and medium transmission range.

The second trace, the MIT cell tower, is used according to the similar principle as that of the IBM trace. The only difference is that instead of functioning as access points, cell towers are used to gather the contact times of the nodes with each other; thus, the resulting network can be characterized as a very dense network due to the high range of the cell tower. The MIT cell tower provides continuous data; therefore, it consists of a large number of contacts with small duration

(less than 10 seconds). Due to several lapses in data gathering, mentioned by the creators of the data, only 89 of 100 devices have been included that visit 32768 different cell towers. Similar to Bluetooth traces, Nov 2004 turned out to be the maximum activity month with 81 devices.

It is imperative to mention that the assumption that two devices connected to one base station (access point or cell tower) introduces inaccuracies [26]. On one hand, it is overly optimistic, since two devices attached to the same access point may still be out of range of each other. On the other hand, the data might omit connection opportunities, since two nodes may pass each other at a place where there is no base station, and this contact would not be logged. Another issue with these datasets is that the devices are not necessarily colocated with their owner at all times (i.e., they do not always characterize human mobility). Despite these inaccuracies, such traces are a valuable source of data, since they span many months and include thousands of devices. Additionally, the datasets used in this study promise to represent the real-life mobility patterns and social networking behaviors of users because the traces are extracted using the mobile devices [27]. Authors in [28] analyzed IBM traces and extracted the usage (session duration and traffic volume) and mobility patterns (number of associated users) of WLAN users. Bhaumik and Batabyal utilized graph tools including centrality and clustering coefficient to propose message dissemination protocols for delay tolerant networks using MIT traces [29]. Further details of the routing simulation mechanism are available in [2].

Centrality computation: as stated earlier, the first two weeks are used for bootstrapping the centrality measures. This process is continued in the latter part of the simulation, and the centrality of each device is updated at a 10-minute interval. This interval is extended to 360 minutes when the devices report their last activity for one day and the next activity occurs on the next morning. The results of the computations are assumed to be transmitted to all devices instantaneously.

Link sharing: each device can maintain the communication session with one other device at any point in time. There are enough independent channels available that any number of node pairs can communicate at the same time with full bandwidth, independent of their proximity to another pair. This aspect will play a key role in analyzing the effect of traffic congestion on devices when *average contact duration* and *intercontact time* are used for centrality measure computation.

Congestion awareness: wireless devices rely on nonsharing channel allocation for data transfer. This aspect is focused to adapt congestion aware metrics based on the amount of data being forwarded through a device. Whenever two devices come into transmission range, messages may be exchanged depending on the current ranking of the devices obtained using one of the centrality measures. When two devices start exchanging messages, then these devices will be invisible to the rest of the surrounding devices. Thus, the ranking obtained using the duration dependent metrics, i.e., *average contact duration* and *intercontact time*, will deteriorate for the devices that attempt to transmit large data.

Table 1: Simulation parameters.

Message count	100
Message size	1.6E3…1.6E7 B
Message lifetime	7 days
Centrality computation interval	10-360 min
Message size distribution	Power law
Bandwidth (low)	100 kB/s
Bandwidth (high)	10,000 kB/s

The peripheral simulation parameters are summarized in Table 1. 100 messages of varying sizes ranging from 1600 B to 1.6E7 B are simulated. The size distribution followed a power law, i.e., a few messages having a large size and many small size messages. We have performed experiments with three centrality measures using the transformed networks. A centralized version of centrality metrics is considered for the sake of comparison. One may consider that the accuracy of metrics will decrease when egocentric variants of the centrality measures based on local information will be available to the individual nodes. A summary of the metrics used with various kinds of centrality is presented in Table 2.

The simulated protocol follows a hop-based routing where every node forwards a message replica to the next node if the centrality measure value of the next node is higher than the current node. In other words, if the receiving node is relatively central to the current node then the current node forwards a replica of one message to another node. The message is replicated during this process and is delivered to the destination if any of the nodes currently in possession of the message replica make a contact with the destination. If the source node of the message has a lower centrality, then the message will be replicated more as compared to the message whose source has a higher centrality in the network.

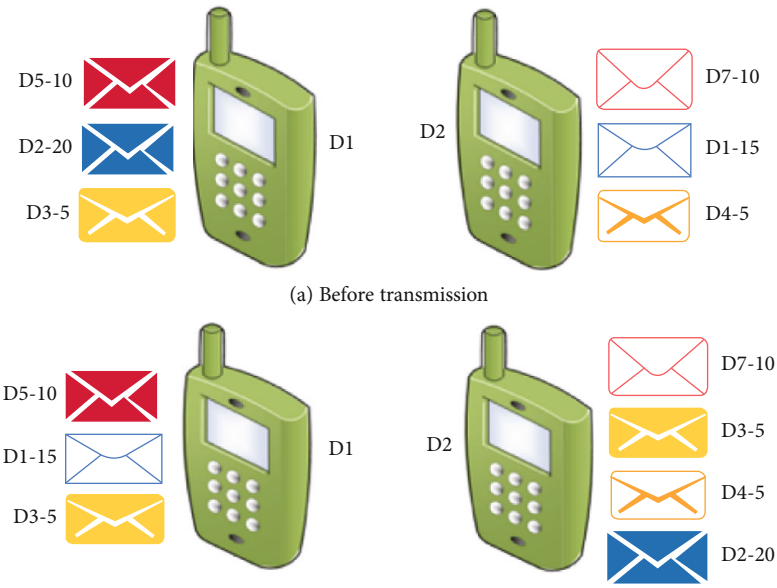
Considering the example presented in Figure 3, two mobile devices D1 and D2 are shown within transmission range before the exchange of messages. Size and destination of each message are shown along each message. It is assumed that the centrality measure of D2 is higher than that of D1. When the transmission phase is over due to the termination of the contact, we can see that two messages destined for each device have been transmitted first followed by the messages that are required to be forwarded. D2 has one more message labelled with blue color that is destined for device D2. However, device D2 that is assumed to have higher centrality value receives two new messages from device D1. One of the messages is destined for device D2, and the other is replicated for forwarding purpose. Two messages destined for D5 and D7 have not been exchanged due to the termination of the contact.

5. Results and Discussions

To establish the correlation between centrality measures and routing importance of nodes, we have simulated epidemic routing to obtain the routing importance of network nodes and then analyzed its relation with the values obtained using centrality measures. A node may have participated in the

Table 2: Transformed metrics and centrality measure combinations.

Transformed metric	Betweenness (B)	Closeness (C)	Degree (D)
Aggregated network (Agr)	X	X	X
Contact count (ContCnt)	X	X	
Contact duration (Dur)	X		
Intercontact duration (ICDur)	X	X	



(a) Before transmission

(b) After transmission

Figure 3: Example showing the working of the message forwarding protocol.

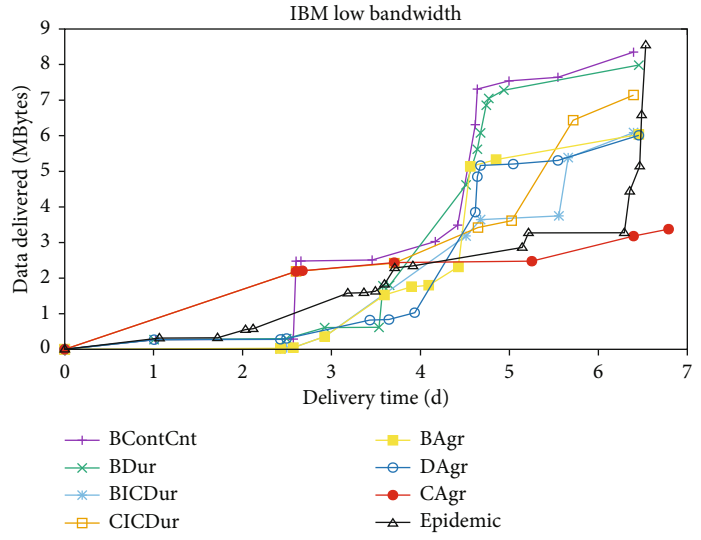
	BContCnt	BDur	BICDur	CContCnt	CICDur	BAgr	DAgr	CAgr	MTCnt
BContCnt	1								
BDur	0.45	1							
BICDur	0.29	0.56	1						
CContCnt	−0.3	−0.42	−0.4	1					
CICDur	−0.32	−0.48	−0.47	0.96	1				
BAgr	0.43	0.72	0.58	−0.47	−0.49	1			
DAgr	0.51	0.74	0.6	−0.74	−0.76	0.81	1		
CAgr	−0.3	−0.41	−0.39	0.99	0.96	−0.46	−0.73	1	
MTCnt	0.49	0.51	−0.42	−0.51	−0.53	−0.61	0.72	−0.51	1

(a) Correlation grid for IBM trace

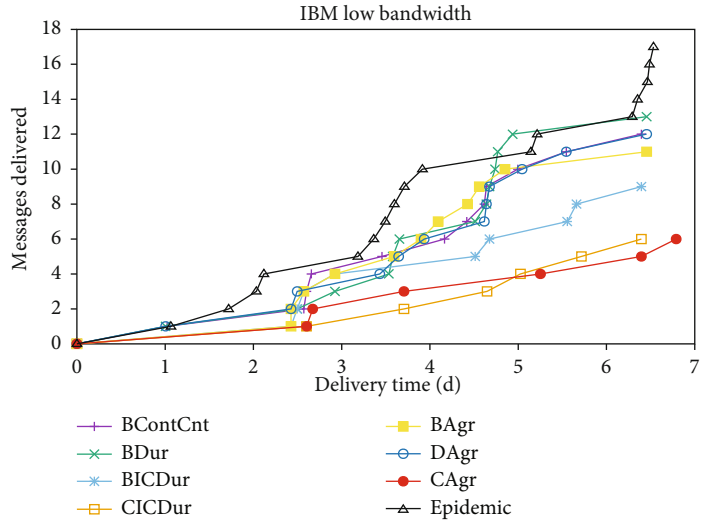
	BContCnt	BDur	BICDur	CContCnt	CICDur	BAgr	DAgr	CAgr	MTCnt
BContCnt	1								
BDur	0.68	1							
BICDur	0.45	0.63	1						
CContCnt	−0.24	−0.27	−0.26	1					
CICDur	−0.46	−0.49	−0.47	0.67	1				
BAgr	0.75	0.77	0.66	−0.26	−0.47	1			
DAgr	0.7	0.78	0.67	−0.51	−0.72	0.75	1		
CAgr	−0.24	−0.27	−0.26	0.99	0.67	−0.26	−0.51	1	
MTCnt	0.69	0.69	−0.66	−0.27	−0.48	−0.78	0.74	−0.26	1

(b) Correlation grid for MITBT trace

Figure 4: Correlation grid for three traces.



(a) Amount of data delivered



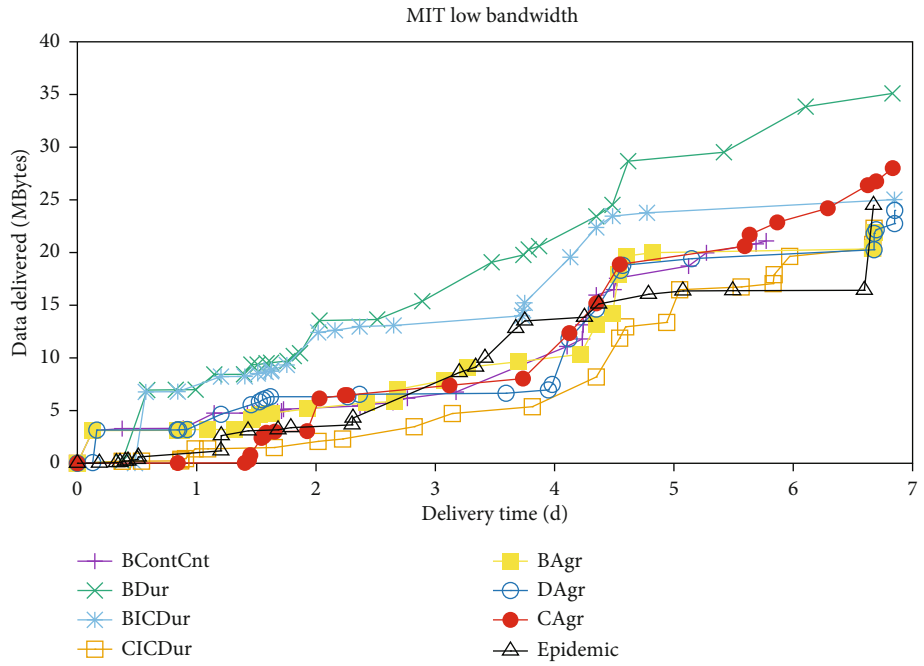
(b) Number of messages delivered

Figure 5: Routing performance of centrality metric IBM network.

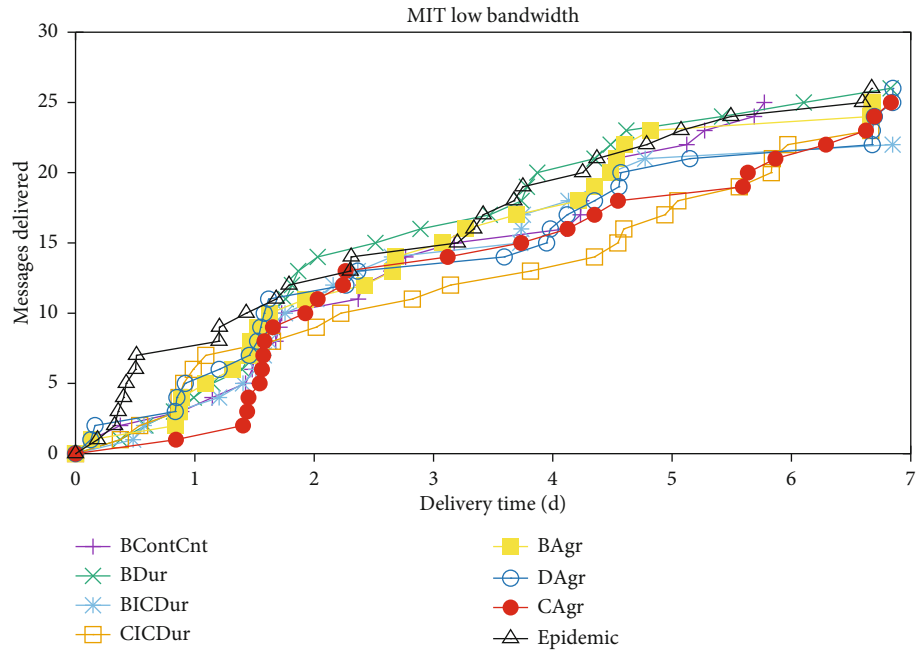
transmission of multiple messages, and we can associate the number of messages, and a node has transmitted to the next-hop with its routing importance. A node that is centrally positioned in the network is expected to participate in the forwarding of a larger number of messages as compared to other nodes that do not have the central position. Correlation is represented among centralities with varying transformed metrics in the form of a grid in Figure 4. The green color represents the positive correlation, and the yellow color shows the negative correlation. The first letter of each label in Figure 4 represents the type of centrality followed by the metric used for centrality computation, i.e., BContCnt represents contact count betweenness. The scheme of abbreviations used in Figure 4 is described in Table 2.

Correlation among investigated centralities is higher in IBM and MITBT traces as compared to the MIT trace. The reason is that MIT consists of a large number of very small duration contacts. As the MIT trace is gathered with the help of cell towers and in many cases, the connection between nodes and cell towers breaks frequently particularly for those nodes that have to select among multiple cell towers due to their location in the overlap area of these cell towers. These aspects result in a high contact count of the MIT trace very high without significantly affecting the duration features of the contacts.

Another aspect observed in Figure 4 is that the same pair of centrality measures for all metrics generally shows a higher correlation with the exception of the MIT trace. Degree centrality shows a consistently negative correlation to the closeness centrality in all the traces. Closeness centrality with respect to intercontact time and contact count shows a negative correlation to messages transmitted. Betweenness centrality with respect to contact count and contact duration shows a positive correlation to the messages transmitted in all traces.



(a) Amount of data delivered

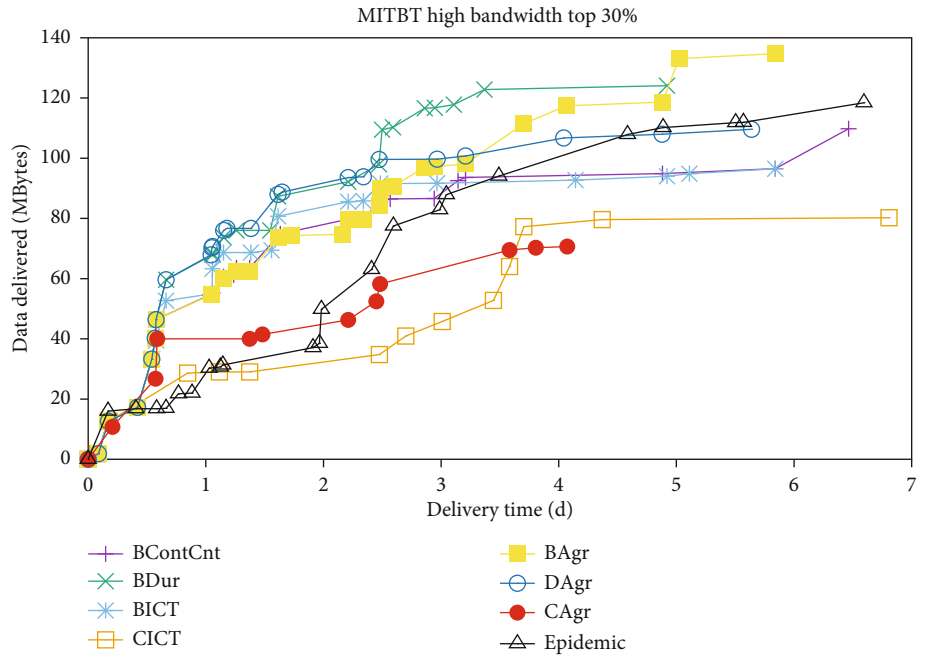


(b) Number of messages delivered

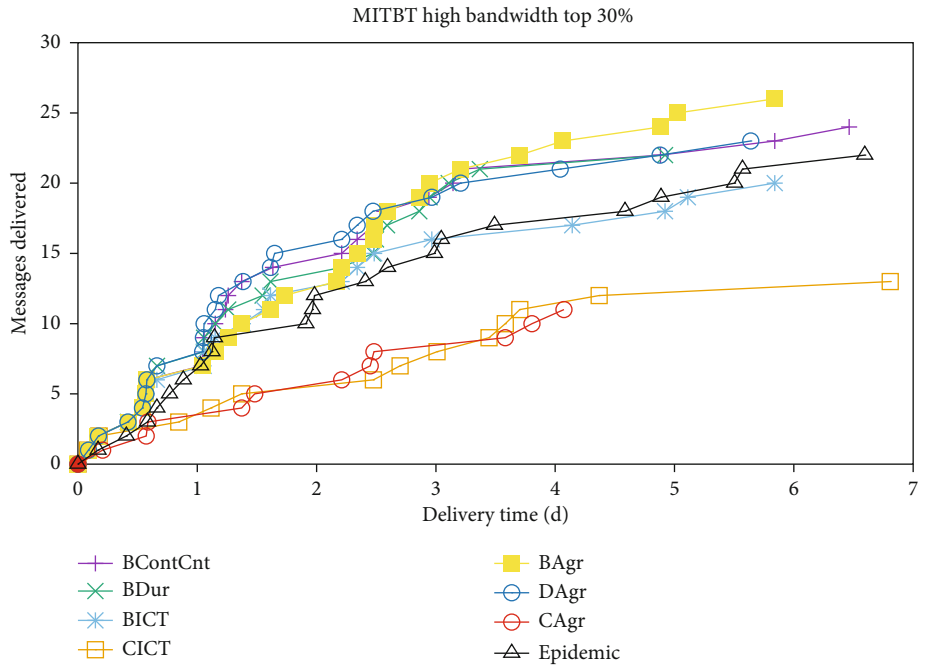
Figure 6: Routing performance of centrality metric MIT network.

The correlation analysis has been used to reduce the scope of the experiments. The simulations are conducted using those combination of centrality measures and metrics that have low correlation with other combinations. From the above discussion, we conclude that the contact count and aggregate network betweenness can be considered as reliable routing metrics for opportunistic network routing for all three traces. Moreover, the correlation shown in Figure 4 is based on the results for the whole trace period, i.e., 1 month.

The results discussed in this section represent several messages and the amount of data delivered during the allocated 7 days of a time span to each message. Each pair of plots consists of the amount of data delivered (a) and messages (b) in each pair of Figures 5–7. For the sake of comparison, we have included epidemic protocol [13] where nodes try to replicate all the messages to the nodes that come in contact with it. An interesting aspect observed in the results of all traces (Figures 5–7) is that majority of centrality metrics have delivered somewhat similar performance to



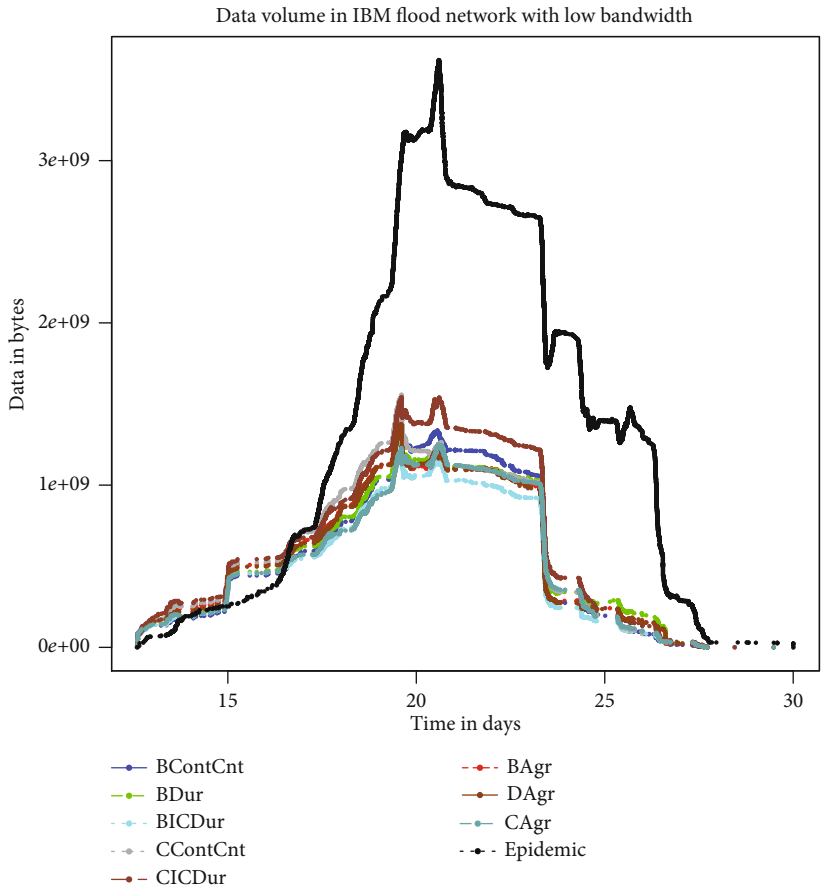
(a) Amount of data delivered



(b) Number of messages delivered

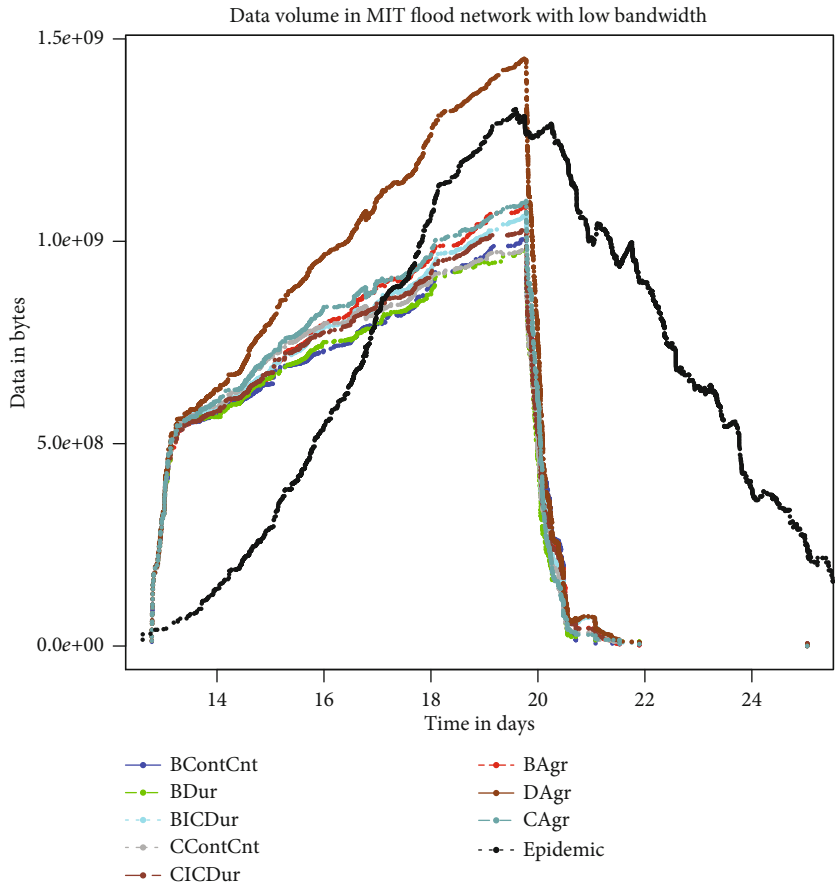
Figure 7: Routing performance of centrality metric MIT network.

epidemic routing. It is imperative to mention that epidemic routing has not been able to attain the best performance in low bandwidth scenarios because devices are not able to forward and successfully replicate messages to other devices due to traffic congestion. A device is not able to forward a message unless it has received a complete replica of the message. The variation in the performance behavior of the centrality metrics in the three traces is due to the variation in contact patterns of the three traces. In the case of MIT trace (small duration frequent contacts) low bandwidth, several of the centrality metrics have performed better than epidemic routing because of the overhead suffered as shown in Figure 6. Small contact durations made epidemic routing performance partially vulnerable as nodes have failed to replicate their messages despite consuming scarce bandwidth, and duration betweenness and contact count betweenness are among the metrics that have delivered the maximum amount of the bytes whereas degree closeness and intercontact time closeness are among the metrics with the low performance.



(a) Data volume for IBM trace

Figure 8: Continued.



(b) Data volume for MIT trace

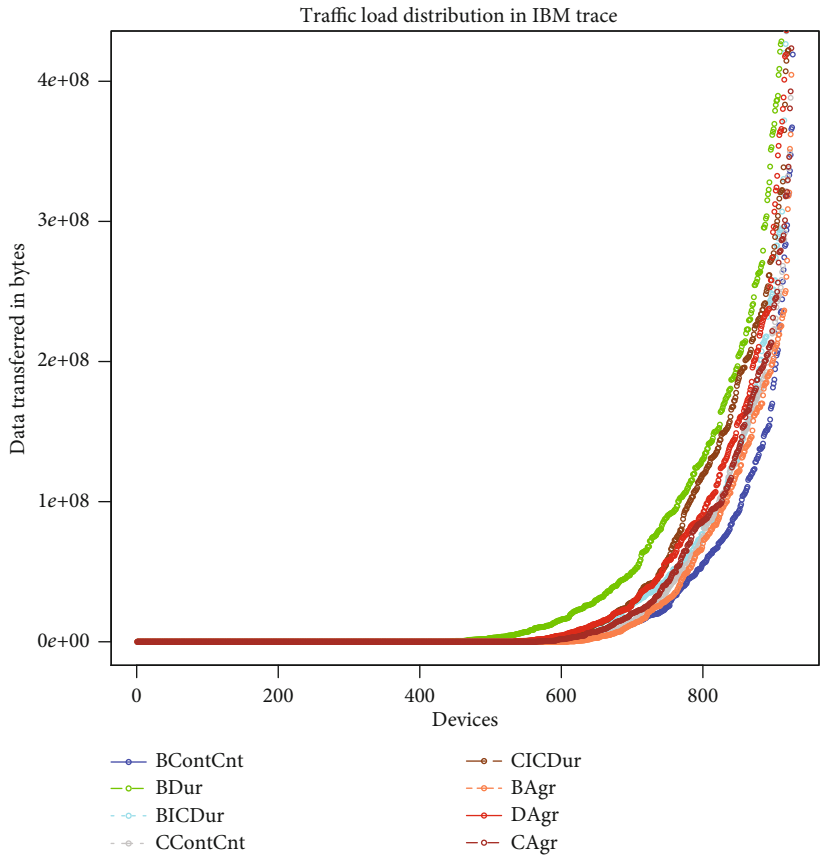
Figure 8: Network memory utilization for IBM and MIT with low bandwidth.

Epidemic routing delivered the maximum number of messages for IBM trace that comprises of contacts that occur with relatively low frequency as shown in Figure 5. It is followed by degree betweenness, contact count betweenness, and contact duration betweenness. In the case of bytes delivered, the list of top performers includes contact count betweenness and contact duration betweenness that have delivered approximately the same number of bytes as epidemic routing, however, in a shorter period. The delivery ratio is the minimum for MITBT trace because it is the sparsest dataset (very low frequency of contacts) among the three that have been utilized for experimentation as shown in Figure 7. The maximum amount of data is delivered by epidemic routing; however, degree betweenness and contact count betweenness have outperformed epidemic routing in several cases. The results are somewhat consistent with IBM and MIT as betweenness-based forwarding mechanism is among the best performance for MITBT as well.

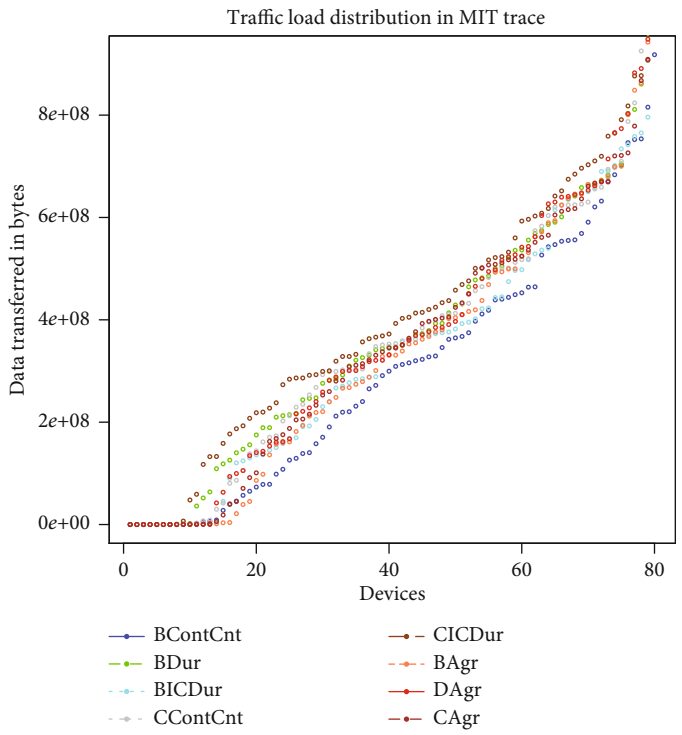
Figure 8 shows the amount of memory consumed in the network, i.e., the consumed network storage increases as the nodes replicate the messages in their possession and transmit to the other nodes by increasing the storage utilization. Once the message is delivered or the lifetime of the message expires, the nodes remove the replica of the message by releasing the local storage. We have assumed the unlimited amount of local storage for each device so that protocols may exploit maximum storage to show maximum performance potential. Taking a closer look at the overheads involved during the centrality metric-based routing protocols as shown in Figure 6, it is noteworthy that protocols using congestion sensitive metrics with betweenness centrality have delivered the competitive number of messages. Also, the respective traffic volume is lower as compared to epidemic routing that shows the utilization of fewer resources.

5.1. Congestion Awareness. When we analyze the routing performance results of Figures 5–7 from the viewpoint of congestion management, the IBM trace shows that contact count betweenness is among the top three centrality forwarding strategies concerning being message count and data volume. As discussed earlier, IBM trace has of relatively lower contact frequency with long durations. The contact-count-based centrality measures have a relatively strong correlation with their aggregate-based counterparts. Aggregate-based centrality measures reported a strong correlation with the message transmission count of each device showing that these measures are prone to be affected by congestion.

Figure 9 shows the load shared by each device during the routing process of all the metrics in IBM and MIT traces. The x-axis represents the devices in sorted order concerning bytes transferred, and y-axis represents the corresponding number of bytes. Both traces (IBM Figure 9(a) and MIT Figure 9(b))



(a) The traffic load on devices in IBM trace



(b) The traffic load on devices in MIT trace

Figure 9: Bytes transferred through individual devices with low bandwidth.

show that average duration and intercontact time have utilized more devices in a relatively balanced way as compared to aggregate and contact count metrics. BDur in IBM trace used more than 450 devices; however, BContCnt used approximately 300 devices, which shows BContCnt burdened a smaller group of devices creating higher congestion than BDur. In the case of MIT trace, both BDur and CICDur have utilized devices in a more balanced way as compared to BContCnt.

6. Conclusion

In this study, several adapted centrality-based routing metrics are evaluated for opportunistic network routing. The centrality measures have been computed using three metrics that preserve the link characteristics among nodes. Influential nodes concerning each centrality measures are identified to analyze the performance of centrality measures. The results show that betweenness centrality-based metrics are twice as good as the closeness centrality metrics. Moreover, the performance of betweenness metrics has been comparable to the epidemic routing. The overhead of all centrality-based routing mechanism is significantly lower than that of epidemic routing.

All transformations can be calculated locally (no global network knowledge required). However, the centrality measure computation has to be adapted to allow any node to estimate its centrality along with its neighbors to make efficient forwarding decisions. In the future, we intend to devise a mechanism to estimate local centrality measures so that the individual nodes can make routing decisions with the help of the information available in their immediate neighborhood [30].

Data Availability

The data used to support the findings of this study have been deposited in the reference [24] repository.

Additional Points

Highlights. The routing simulations in this paper are performed by adapting the real-life social networks that are extracted using the traces of wireless devices. The results presented in this paper are extended from [31] which indicates that the message delivery ratio of the congestion aware metrics is observed compatible with the message delivery ratio for epidemic routing. The highlights of the article are as follows: evaluation of routing performance of three opportunistic network metrics, i.e., *contact count*, *intercontact time*, and *average contact duration*, to harness link congestion information; a comprehensive congestion analysis concerning both complete network and individual nodes of the network metrics by simulation using three real-life social networks.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Dhungana and E. Bulut, "Energy balancing in mobile opportunistic networks with wireless charging: single and multi-hop approaches," *Ad Hoc Networks*, vol. 111, article 102342, 2021.

[2] M. Islam and M. Waldvogel, "Optimizing message delivery in mobile-opportunistic networks," in *Baltic Congress on Future Internet Communications*, pp. 134–141, Riga, Latvia, 2011.

[3] A. Jain, "Betweenness centrality based connectivity aware routing algorithm for prolonging network lifetime in wireless sensor networks," *Wireless Networks*, vol. 22, no. 5, pp. 1605–1624, 2016.

[4] Y. Xiao and J. Wu, "Data transmission and management based on node communication in opportunistic social networks," *Symmetry*, vol. 12, no. 8, p. 1288, 2020.

[5] L. C. Freeman, "Centrality in social networks: conceptual clarification," *Social Networks*, vol. 1, pp. 215–239, 1979.

[6] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

[7] P. Bonacich, "Power and centrality: a family of measures," *American Journal of Sociology*, vol. 92, no. 5, pp. 1170–1182, 1987.

[8] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[9] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[10] C. Miralda, E. Donald, M. Ikeda, K. Mutso, L. Barolli, and M. Takizawa, "Effect of node centrality for IoT device selection in opportunistic networks: a comparison study," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 21, article E4790, 2018.

[11] A. Rasheed, S. Amjal, and A. Qayyum, "Adaptive routing update approach for VANET using local neighbourhood change information," *Malaysian Journal of Computer Science*, vol. 27, pp. 307–327, 2014.

[12] P. Yuan, P. Liu, and S. Tang, "Exploiting partial centrality of nodes for data forwarding in mobile opportunistic networks," in *IEEE 17th International Conference on Computational Science and Engineering (CSE)*, Chengdu, China, 2014.

[13] G. Goudar and S. Batabyal, "Point of congestion in large buffer mobile opportunistic network," *IEEE Communications Letters*, vol. 24, no. 7, pp. 1586–1590, 2020.

[14] H. B. Liaqat, F. Xia, Q. Yang, L. Liu, Z. Chen, and T. Qiu, "Reliable TCP for popular data in socially-aware ad-hoc networks," in *Proceedings of the 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, New York, NY, USA, 2014.

[15] J. Wang, C. Jiang, T. Q. S. Quek, X. Wang, and Y. Ren, "The value strength aided information diffusion in socially-aware mobile networks," *IEEE Access*, vol. 4, pp. 3907–3919, 2016.

[16] H. Zhu, M. Dong, S. Chang, Y. Zhu, M. Li, and X. Shen, "Zoom: scaling the mobility for fast opportunistic forwarding in vehicular networks," in *IEEE INFOCOM Proceedings*, pp. 2832–2840, Turin, Italy, 2013.

[17] P. Santi, *Mobility Models for Next Generation Wireless Networks*, John Wiley & Sons Ltd, 2012.

[18] M. Tulu, M. Mkiramweni, R. Hou, S. Feisso, and T. Younas, "Influential nodes selection to enhance data dissemination in

mobile social networks: a survey," *Journal of Network and Computer Applications*, vol. 169, article 102768, 2020.

[19] C. Miralda, E. Donald, M. Ikeda, K. Mutso, L. Barolli, and M. Takizawa, "A decision-making system based on fuzzy logic for IoT node selection in opportunistic networks considering node betweenness centrality as a new parameter," in *International Conference on Intelligent Networking and Collaborative Systems*, Cham, 2020.

[20] C. M. Kim, I. S. Kang, Y. H. Han, and Y. S. Jeong, "An efficient routing scheme based on social relations in delay-tolerant networks," in *Lecture Notes in Electrical Engineering Heidelberg*, pp. 533–540, Springer, Berlin, 2014.

[21] K. M. Sivalingam and V. Chellappan, "Application of entropy of centrality measures to routing in tactical wireless networks," in *2013 19th IEEE Workshop on Local Metropolitan Area Networks (LANMAN)*, Brussels, Belgium, 2013.

[22] T. Amah, M. Kamat, K. Bakar, W. Moreira, A. Oliveira, and M. Batistita, "Preparing opportunistic networks for smart cities: collecting sensed data with minimal knowledge," *Journal of Parallel and Distributed Computing*, vol. 135, pp. 21–55, 2020.

[23] R. Akbar, F. Safaeia, and E. Khodadad, "A novel adaptive congestion-aware and load-balanced routing algorithm in networks-on-chip," *Computers and Electrical Engineering*, vol. 71, pp. 60–76, 2018.

[24] N. P. A. Eagle, "Reality mining: sensing complex social systems," *Journal Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, 2006.

[25] P. Cao, G. Li, A. Champion, D. Xuan, S. Romig, and W. Steve, "On human mobility predictability via WLAN logs," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, Atlanta, GA, USA, 2017.

[26] A. Trivedi, J. Gummeson, and P. Shenoy, "Empirical characterization of mobility of multi-device internet users," 2020, https://arxiv.org/abs/2003.08512.

[27] G. Poucin, B. Farooq, and Z. Patterson, ""Activity patterns mining in Wi-Fi access point logs," Computers," *Envoirnment and Urban Systems*, vol. 67, pp. 55–67, 2018.

[28] F. Ganji, Ł. Budzisz, F. G. Debele et al., "Greening campus WLANs: energy-relevant usage and mobility patterns," *Computer Networks*, vol. 78, pp. 164–181, 2015.

[29] P. B. Suvadip Batabyal, "Analysing social behaviour and message dissemination in human based delay tolerant network," *Wireless Networks*, vol. 21, no. 2, pp. 513–529, 2015.

[30] Z. Gao, Y. Shi, and S. Chen, "Identifying influential nodes for efficient routing in opportunistic networks," *Journal of Communications*, vol. 10, no. 1, pp. 48–54, 2015.

[31] M. A. Islam, M. A. Iqbal, M. Aleem, and Z. Halim, "Analysing connectivity patterns and centrality metrics for opportunistic networks," in *International Conference on Communication, Computing and Digital Systems (C-CODE)*, Islamabad, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9962697, 10 pages
https://doi.org/10.1155/2021/9962697

Research Article

A Novel QoS-Oriented Intrusion Detection Mechanism for IoT Applications

Abdulfattah Noorwali,[1] **Ahmad Naseem Alvi**,[2] **Mohammad Zubair Khan**,[3] **Muhammad Awais Javed**,[2] **Wadii Boulila**,[4] **and Priyadarshini A. Pattanaik**[5]

[1]*Department of Electrical Engineering, Umm Al-Qura University, Makkah 21961, Saudi Arabia*
[2]*Department of Electrical and Computer Engineering, COMSATS University Islamabad, 45550, Pakistan*
[3]*Department of Computer Science, College of Computer Science and Engineering, Taibah University, Medina, Saudi Arabia*
[4]*RIADI Laboratory, National School of Computer Science, University of Manouba, Tunisia*
[5]*Image and Information Processing Department, IMT Atlantique, LaTIM Inserm U1101, Brest 29238, France*

Correspondence should be addressed to Wadii Boulila; wadii.boulila@riadi.rnu.tn

Received 3 April 2021; Accepted 4 June 2021; Published 18 June 2021

Academic Editor: Nawab Muhammad Faseeh Qureshi

Wireless sensor network (WSN) is an integral part of Internet of Things (IoT). The sensor nodes in WSN generate large sensing data which is disseminated to intelligent servers using multiple wireless networks. This large data is prone to attacks from malicious nodes which become part of the network, and it is difficult to find these adversaries. The work in this paper presents a mechanism to detect adversaries for the IEEE 802.15.4 standard which is a central medium access protocol used in WSN-based IoT applications. The collisions and exhaustion attacks are detected based on a soft decision-based algorithm. In case the QoS of the network is compromised due to large data traffic, the proposed protocol adaptively varies the duty cycle of the IEEE 802.15.4. Simulation results show that the proposed intrusion detection and adaptive duty cycle algorithm improves the energy efficiency of a WSN with a reduced network delay.

1. Introduction

Internet of Things (IoT) applications use wireless sensor networks (WSNs) to implement several applications in the fields of healthcare, military, environmental monitoring, smart cities, agricultural engineering, etc. WSNs collect sensing data from different areas, generate large data sets, and share it with remote servers for intelligent processing and valuable insights. The data dissemination for IoT applications is shared locally using sensor nodes, then passed to the nearby computing nodes, and finally reached the centralized server, thus forming a reliable multihoming network. WSNs comprise tiny wireless nodes that operate autonomously on a battery with limited energy, so they are required to be energy efficient. Besides, sensor nodes have limited computational capabilities along with low data rates and low processing.

At the data link layer, many medium access control (MAC) algorithms have been proposed in literature that consider these limitations to meet application requirements. The main concerns in these MAC protocols include energy efficiency, delay minimization, and better throughput. These MAC protocols also offer nodes scalability, reliability, and adaptability [1–3].

IEEE 802.15.4 standard was developed to transmit data using low transmit power and low data rate to nearby nodes. It is widely accepted and embedded in the majority of the WSNs [4]. The standard has two operating modes: first is the beacon-enabled mode and second is the non-beacon-enabled mode. The first mode, i.e., beacon-enabled mode, is mostly used for IoT applications. Figure 1 shows the superframe diagram of IEEE 802.15.4. The duty cycle of IEEE 802.15.4 standard is controlled using two variables: first is

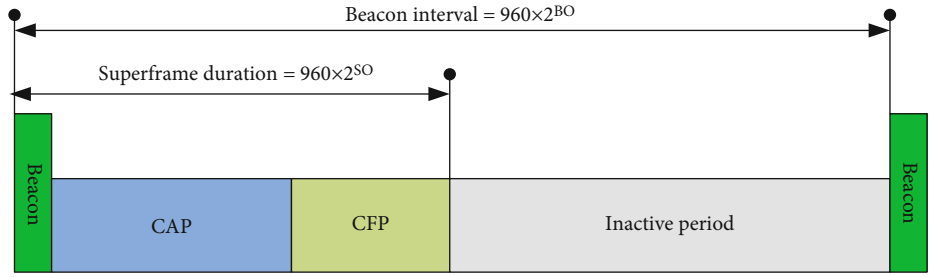


Figure 1: IEEE 802.15.4 standard superframe.

the beacon order (BO) and second is the superframe order (SO). A higher value of BO increases the duration of the beacon known as beacon interval (BI). The active portion of the superframe is controlled by SO. The superframe structure can work in a hybrid manner, using both contention-free access mechanism and also contention-based access technique. Slotted carrier sense multiple access (CSMA/CA) is used for coordinating multiple access during contention-based period. On the other hand, guaranteed time slots are allocated for data transmission without contention.

Most WSN applications use the IEEE 802.15.4 standard due to its hybrid capabilities. Some applications are sensitive in nature, and vulnerability in data delivery is unacceptable [5–7]. Due to its wide utility, the beacon-enabled mode of the standard remains in prime research. Its contention-based and contention-less modes are examined and evaluated in different prospects, such as efficient slot allocation for better link utilization [8–10].

Security issues are the main concern in sensitive WSN and IoT applications such as remote healthcare, traffic safety, and home security; that is why many security measures have been proposed [11, 12]. Most of the solutions are related to encryption to ensure that the intruders should not understand the data. For this purpose, cryptographic techniques are widely used to ensure security of IoT applications and provide data integrity, confidentiality, and privacy [13]. Adversaries try to disturb a wireless network to disrupt communications, especially in highly sensitive applications. Most of these attacks target physical and MAC layers. That is why the IEEE 802.15.4 standard remains a high target point of these adversaries, especially during the contention-based or contention-less periods. These attacks not only reduce the network efficiency of the standard but also decrease the network lifetime.

There are several techniques in the literature to determine the intruder's attacks in different MAC protocols of WSNs. Yu et al. [14] provide a solution to detect spoofing attacks by examining network characteristics. Intrusion attacks that do not occur often are hard to detect, and authors in [15] propose a technique to detect these attacks efficiently. Authors in [16, 17] introduce mechanisms to detect continuous wave jamming signals in the IEEE 802.15.4 standard. Besides these, many other recent techniques focused on intrusion detection [18–20]. Security and intrusion detection is a key component of future IoT applications and 6G technologies [21–23].

Most of the research about intruder's attack detection is based on jamming and spoofing attacks and does not identify the intruder's attacks causing collisions and packet exhaustion. In this paper, an intrusion detection technique is proposed to detect adversaries' attacks in the IEEE 802.15.4. Besides, an adaptive duty cycle algorithm is proposed to meet the QoS. The salient features of our proposed work include the following:

(1) A soft function is developed to compute the probabilities of the collision, successful transmission, and packet exhaustion

(2) An intrusion detection mechanism is developed to detect the collision and exhaustion attacks from this soft function

(3) An adaptive duty cycle algorithm is proposed that works for IEEE 801.5.4 standard and achieves required QoS in a variable traffic scenario

The paper starts by describing the working of IEEE 802.15.4 standard in Section 2, followed by an intrusion detection mechanism in Section 3. The network performance with and without intrusion detection is presented in Section 4. The conclusion of the paper is given in Section 5.

2. IEEE 802.15.4 Working

IEEE 802.15.4 standard is designed for applications that work on low transmission powers and data rates. The standard can work on three frequency bands: first is the 868 MHz, second is the 915 MHz, and last one is the 2.4 GHz. The first two spectrum bands do not need license and offer 1 and 10 frequency channels, respectively. 868 and 915 MHz use the BPSK modulation scheme and data rates of 20,000 and 40,000 bits/sec, respectively. However, 2.4 GHz comprises 16 different frequency channels and uses an O-QPSK modulation scheme by offering a 250,000 bits/sec data rate. A comparative table of all these frequency bands is given in Table 1.

The standard offers both ad hoc and centralized controlled network. In an ad hoc manner, nodes send their information with each other using an unslotted CSMA/CA-based multiple access algorithm. In a centralized network, a superframe architecture is used that has active and inactive periods. The coordinator issues a beacon frame, and IoT nodes turn on their transceivers to receive the message and

Table 1: Frequency spectrum of IEEE 802.15.4 standard.

Frequency spectrum (MHz)	Modulation scheme	Symbols rate	Symbols rate	Number of channels	Bit duration (sec)
868	BPSK	20k	20k	1	$50 * 10^{-6}$
915	BPSK	40k	40k	10	$25 * 10^{-6}$
2400	O-QPSK	62.5k	250k	16	$4 * 10^{-6}$
868	BPSK	20k	20k	1	$50 * 10^{-6}$

keep them synchronized. The active period known as superframe duration (SD) contains 16 equal duration slots. SD includes transmission of the beacon frame, transmission of data using contention access period (CAP), and data dissemination in the contention-free period (CFP). Out of these 16 time slots, beacon and CAP can use 9 or more slots, whereas CFP can be allocated at most 7 slots.

The coordinator periodically generates beacon frames. Nonmember nodes that intend to become a member of the network must wait for the beacon to know the CAP period to transmit their joining requests to the coordinator. The data requests of IoT nodes are processed during CAP, and the coordinator then allocates guaranteed time slots (GTS) to these IoT nodes. Those nodes that are not allocated GTS can transfer their data during CAP by following the multiple access slotted CSMA/CA protocol. That is why the superframe structure has a mandatory portion of CAP where nodes send their requests and data by applying a slotted CSMA/CA algorithm.

2.1. Slotted CSMA/CA Working. During CAP, the coordinator keeps its radio on to receive and respond to all the requests originated by the nodes. All nodes that intend to send any requests or transmit data during CAP use the slotted CSMA/CA protocol to confirm the idleness of the medium. The slotted CSMA/CA comprises three main parameters, named NB, BE, and CW.

Parameter NB is the maximum number of times a node can check medium availability for transmission. The default value of NB is in the range of 0 to 4, which means the maximum attempts to assess the medium availability is 5 when NB is 4. If a node does not find medium free even after NB, it informs the upper layer about channel access failure.

Parameter BE is backoff exponent, and its default value ranges from 3 to 5. It determines a random wait time that the node takes before checking the medium availability for transmission. The random range of the backoff period is always in the range of 0 to $2^{BE} - 1$. In case the medium is busy, it increments its value by 1 until the maximum value is reached. For an initial value of 3, a node should wait for any random value of backoff periods between 0 and 7. The maximum default backoff period for a node to wait before checking medium availability in this standard is 31 backoff periods.

Parameter CW stands for the contention window with a default value of 2. It decrements by 1 whenever finding the channel idle and reset to its initial value of 2 by finding the channel busy. Nodes can send their data when the value of CW becomes 0. This allows the nodes to confirm the channel idle by applying the clear channel assessment (CCA). CSMA/CA flow diagram of CSMA/CA is presented in Figure 2.

When data is transmitted, then the node waits for the acknowledgment. Suppose it could not receive within a specified time. In that case, the same frame is transmitted again until it exceeds the maximum frame retries limit, which has been defined in a MaxFrameRetries parameter.

Most of the adversaries do not follow the algorithm and do not follow the standard policies, severely disturbing the network quality by transmitting the smaller or longer stream of packets in the medium. A description of different adversary attacks on IEEE 802.15.4 standard is given in the following section.

2.2. Attacks on the IEEE 802.15.4 Standard. This section describes some of the intruder's attacks on the standard, which ultimately affects the WSN performance. The following three attacks are quite common and disturb either CAP, CFP, or both.

(i) Exhaustion attack (CAP)

(ii) Collision attack (CAP+CFP)

(iii) Unfairness attack (CFP)

2.2.1. Exhaustion Attack. In most WSN applications, legitimate nodes are deployed in an open environment. In the IEEE 802.15.4 standard, each member node applies a slotted CSMA/CA algorithm before sending its packet to other nodes or coordinator. Suppose it could not send its packet due to the channel busy after exceeding its threshold limits as described in the algorithm. In that case, it reports the channel access failure to its upper layer, and the coordinator gets this information through an indirect data transfer method. After completing their backoff counter, each requesting node detects the channel availability by performing a clear channel assessment (CCA). Adversaries keep the channel busy by sending a long stream of messages. Due to these attacks, legitimate nodes always find the channel busy in each CCA operation and find channel access failure after exceeding its retry limits. Besides, when a node transmits its packet and could not receive its acknowledgment, it must resend the packet repeatedly till its maximum limit and then finally declares that the packet cannot be transmitted.

2.2.2. Collision Attack. The collision occurs when a node communicates with another node while the third node sends packets on the same frequency channel. In the IEEE 802.15.4

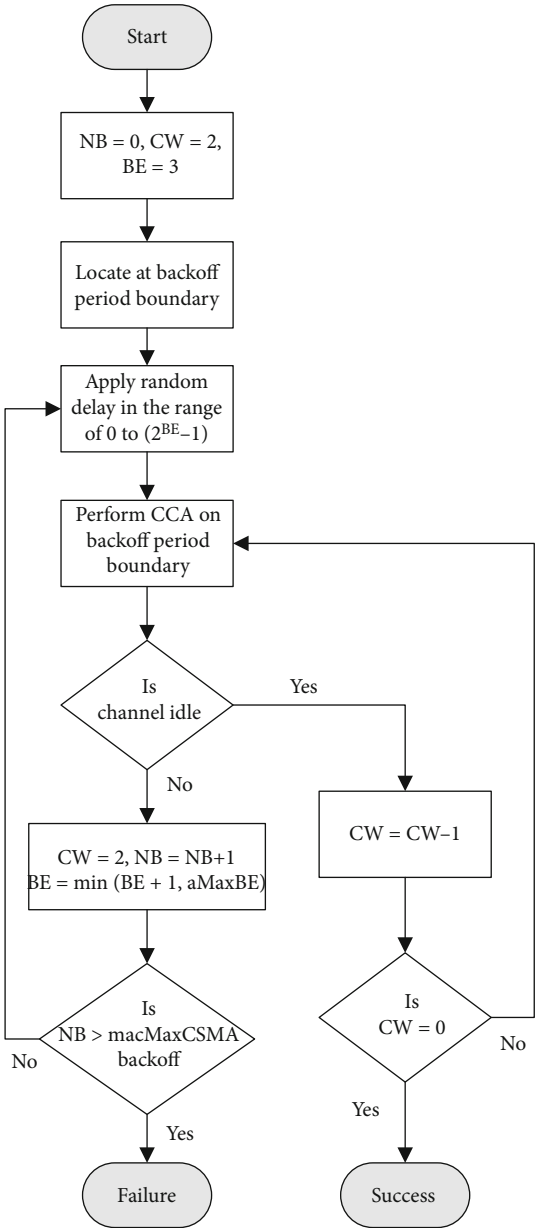


Figure 2: CSMA/CA mechanism of IEEE 802.15.4 standard.

standard, nodes transmit their information with and without contending the medium with other nodes during CAP and CFP, respectively. Adversaries intentionally transmit small packets during these transmissions that cause the collision, and nodes must resend their packets from scratch. This retransmission consumes energy and increases the transmission delay, and ultimately, the quality of service is compromised.

2.2.3. Unfairness Attack. The standard allocates GTS to data requesting nodes on a first-come, first-served basis. All the data requesting nodes have equal chances of generating their request during CAP as they are required to observe the clear channel availability after decrementing the backoff period. The adversary node does not follow the backoff period and sends its data request soon after the beacon frame. The coordinator receives GTS requests from the adversaries before other legitimate requests and allocates GTS before the adversaries compared to the legitimate nodes.

Most of the adversary attacks are during CAP because they cannot be determined easily. In this work, an intrusion detection mechanism is proposed specifically for CAP of the IEEE 802.15.4 standard by identifying collision and exhaustion attacks. Besides, an adaptive duty cycle algorithm is proposed that allows the coordinator to adjust its duty cycle to meet the QoS.

3. Proposed Scheme

This work helps the coordinator to determine the intruder's attack and satisfies the QoS by introducing two different algorithms. The first algorithm checks whether the QoS is compromised due to an intruder's attack or not. In case there are no intruder's attacks, the second algorithm allows the coordinator to adjust its duty cycle to satisfy its QoS.

3.1. Intrusion Detection Mechanism. In this section, a mechanism for intrusion detection in the IEEE 802.15.4 standard is proposed based on the following inputs:

(1) Collision ratio (CR): number of collisions detected against the total packets transmitted by a node per second

(2) Packet successful transmission ratio (PST): ratio of packets received at destination to the packets transmitted by the source node

(3) Request ratio (RR): ratio of the number of requests generated by the nodes to the coordinator's requests per second

A soft function is designed to find out the probability of collision (PC), probability of successful transmission (PS), and probability of exhaustion (PE) from input values of CR, PST, and RR, respectively, as follows:

$$Y(V) = \frac{1}{1 + \exp\{-E * (V - F)\}}. \quad (1)$$

Here, $Y(V)$ gives the probability of collision, successful transmission, and exhaustion from this soft function when we input CR, PST, and RR values by putting these values against V in the soft function. E in this soft function is the slope parameter, and F is the center of the curve.

The shape of the curve depends upon the values of E and F, and their values can be determined by a cost function as follows:

$$J(V) = (Y_D - Y)^2. \quad (2)$$

Here, Y is the actual value and Y_D is the desired value.

Input: Collision Ratio *CR*, Packet Successful Transmission ratio *PST*, Request Ratio *RR*
Calculate $PC = 1/[1 + exp\{-E\times(CR - F)\}]$
Calculate $PS = 1/[1 + exp\{-E\times(PST - F)\}]$
Calculate $PE = 1/[1 + exp\{-E\times(RR - F)\}]$
Calculate $Z_1 = (PS\times\varphi) + (PC\times\theta)$
Calculate $Z_2 = (PS\times\varphi) + (PE\times\theta)$
if $Z_1 > Th$
Collision attack found
else
No Collision attack
if $Z_2 > Th$
Exhaustion attack found
Else
No Exhaustion attack

Algorithm 1: Intruder detection algorithm.

Values of E and F change periodically. If E_N and F_N are their current values, then their next stages E_{N+1} and F_{N+1} can be determined from their current values as follows:

$$E_{N+1} = E_N + \varnothing * \frac{\partial J}{\partial E}, \tag{3}$$

where $\varnothing$ ranges from 0 to 1 and $\partial J/\partial E$ is calculated as follows:

$$\frac{\partial J}{\partial E} = 2(Y_D - Y)\frac{E_N}{[1 + \exp\{-E_N * (V - F_N)\}]^2}. \tag{4}$$

And F_{N+1} is calculated as follows:

$$F_{N+1} = F_N + \varnothing * \frac{\partial J}{\partial F}. \tag{5}$$

Here, $\partial J/\partial F$ is calculated as follows:

$$\frac{\partial J}{\partial F} = 2(Y_D - Y)\frac{-E_N * \exp[-E_N * (V - F_N)]}{[1 + \exp\{-E_N * (V - F_N)\}]^2}. \tag{6}$$

The more the number of iterations to find out the next stages of E and F, the steeper the curve will be.

After finding out the probabilities of collisions, successful transmissions, and exhaustion from input values of CR, PST, and RR, respectively, intruder's attacks can be determined as shown in Algorithm 1. This algorithm detects collisions and exhaustion attacks of the intruders.

3.1.1. Collision Attack Detection. Algorithm after determining collisions and successful transmission probabilities from the soft function scale them by φ and θ, respectively. Both scaled values are summed up and then compared with the threshold value. In case the summed value is greater than the threshold value, an attack is found; else, no attack is found.

3.1.2. Exhaustion Attack Detection. Adversary exhaustion attacks are detected from the probability of exhaustion and successful transmission probability. Like the above criteria, the probability of success is summed individually with the probability of exhaustion and compared with the threshold. The threshold is set, and then, this summation result is compared with the threshold. If the sum is greater than the threshold, an attack is found; else, no attack is found.

3.2. Duty Cycle Adjustment Algorithm. In case the intruder detection algorithm could not detect the adversary's involvement and the system still feels that QoS is compromised, the coordinator's duty cycle needs to be adjusted to satisfy the QoS. This algorithm allows the coordinator to adjust its duty cycle to meet the QoS. In case the member of nodes is increased and requires more data requests, an increase in the active period is achieved. Similarly, when there is less data in a network, then the active period needs to shrink. IEEE 802.15.4 standard does not discuss the adaptation in the duty cycle according to the QoS requirement. This algorithm allows the coordinator to adjust its duty cycle to meet the QoS of the network.

The algorithm improves the data transmission by improving the throughput and reduces the number of collisions. The algorithm allows the coordinator to test QoS during the last BI and make proper adjustments in the next BI. It is expected that the data required to receive by the coordinator is half of its maximum capacity. In case the received data is less than this threshold value, then it may be due to an increased number of collisions, as shown in Figure 3.

If QoS is compromised only due to low throughput, then the algorithm allows the coordinator to adjust the duty cycle as follows:

(i) If the difference between BO and SO is more than 1 and SO is greater than 0, then the BO and SO's parameter value will be reduced in the next BI. This decreases the active duration that allows nodes to send their data in less time, resulting in increased throughput

(ii) In case the difference between BO and SO is greater than 1, however SO is 0, then BO will be reduced without any change in SO. This decrease in BO reduces the inactive period, and consequently, nodes' waiting time is reduced

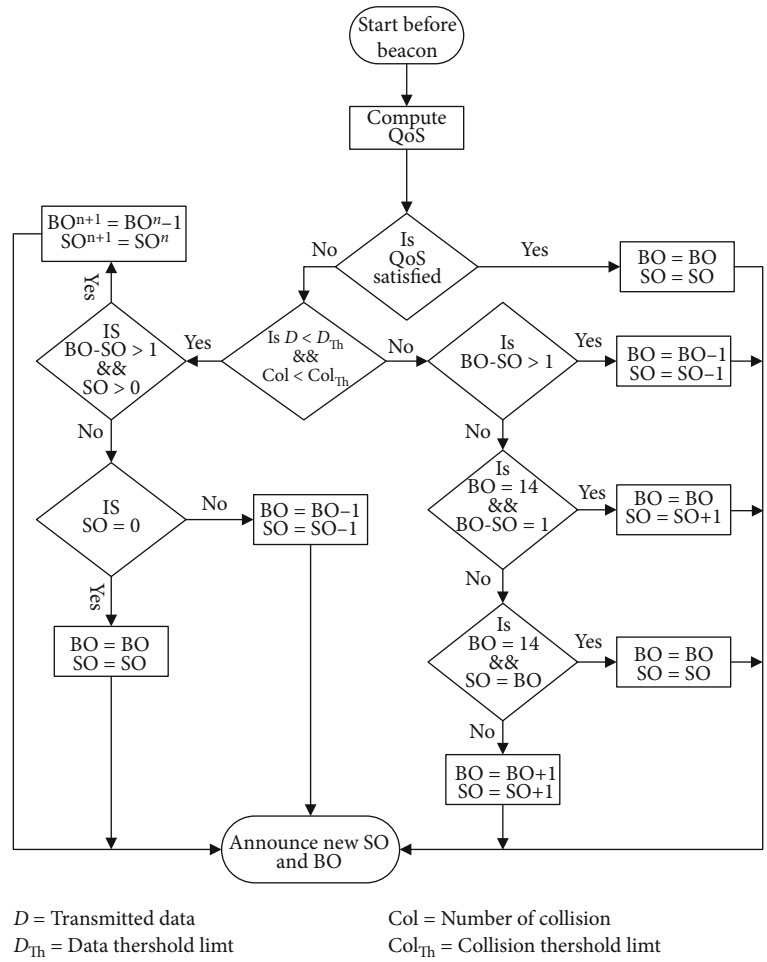


Figure 3: Proposed adaptive duty cycle adjustment.

(iii) If the difference between BO and SO is 0 and SO is greater than 1, then SO's next value is decremented without any change in BO. The decrease in SO results in a reduced active period, and hence, the throughput is increased

(iv) If the difference in BO and SO is 0 and SO's value is 0, then they remain unchanged during the next BI. This is the scenario when data traffic in a network is very small

If there is an increased number of collisions during BI, it may reduce the throughput. This increase in the collision may occur due to insufficient active duration, and there is a requirement to increase the active period so that nodes have more time to send their data requests. Following are the criteria for adjusting the parameter values of SO and BO.

(i) If the difference between BO and SO is greater than 0 and BO is less than 14, then both SO and BO will be increased in the next BI

(ii) In case BO approaches its maximum limit of 14, and the difference between BO and SO is 0, then SO will be increased without any change in BO value

Table 2: Parameter values.

Parameters	Values
Number of nodes	10
Network size	100
Data rate (kbps)	250
Legitimate nodes	15
Intruder nodes	5
Sink node	1
Superframe duration (msec)	122.88
Beacon interval (msec)	245.76
Duty cycle (%)	50
Transmitting energy (mJ)	50
Receiving energy (mJ)	30
Idle energy (mJ)	5
Offered load (kbytes)	1 : 1 : 10

(iii) If both SO and BO are 14, then the next SO and BO value will remain unchanged as it is already on its maximum limit

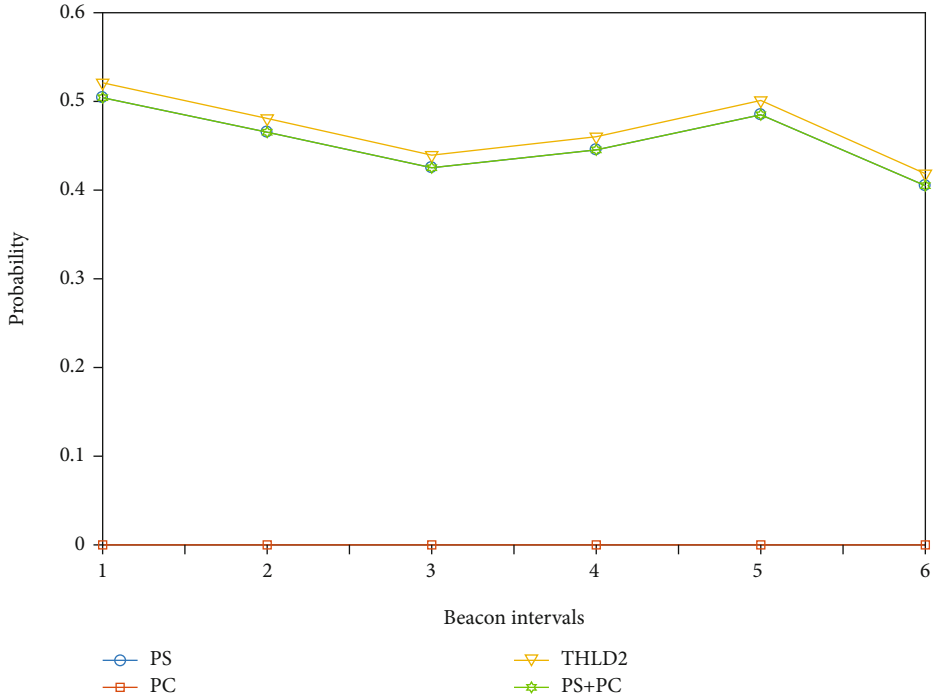


Figure 4: Scenario when no attack found against beacon intervals.

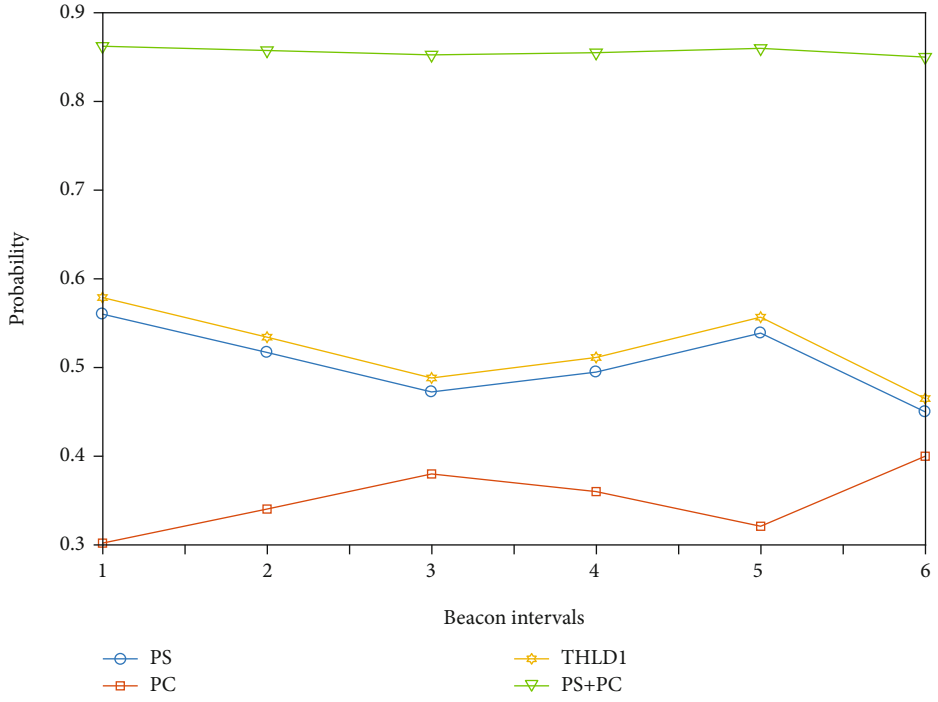


Figure 5: Collision attack detection against beacon intervals.

(iv) If BO is less than 14 with a 100% duty cycle, then we have the flexibility to increase the SO. However, we must increase the value of BO to meet the standard permissible requirements

4. Analysis and Results

In this section, the proposed scheme is evaluated by developing a simulation environment in MATLAB, as shown in

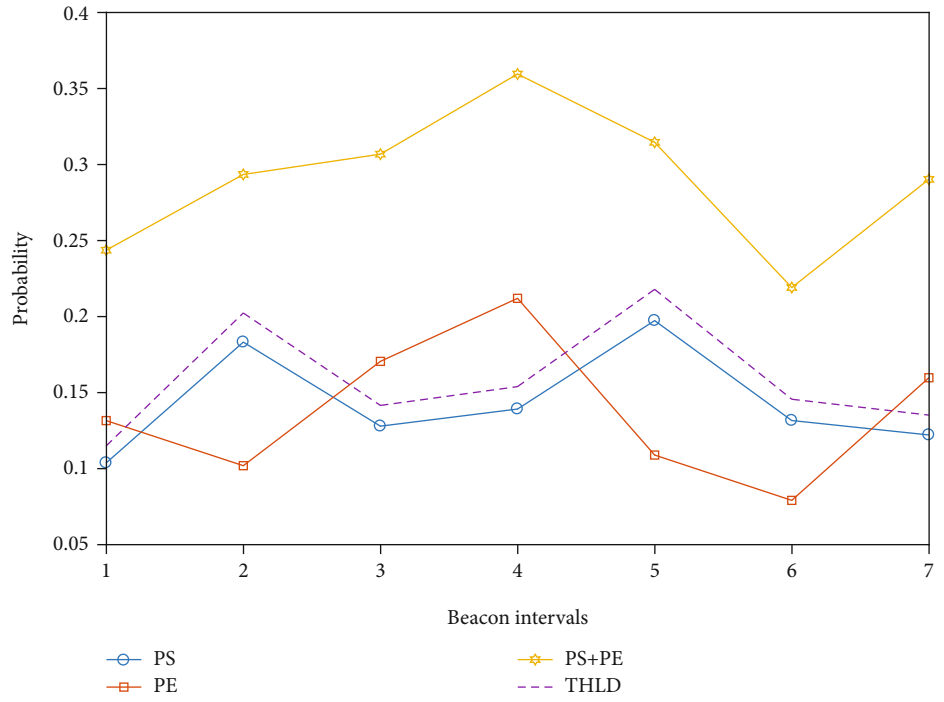


Figure 6: Exhaustion attack detection.

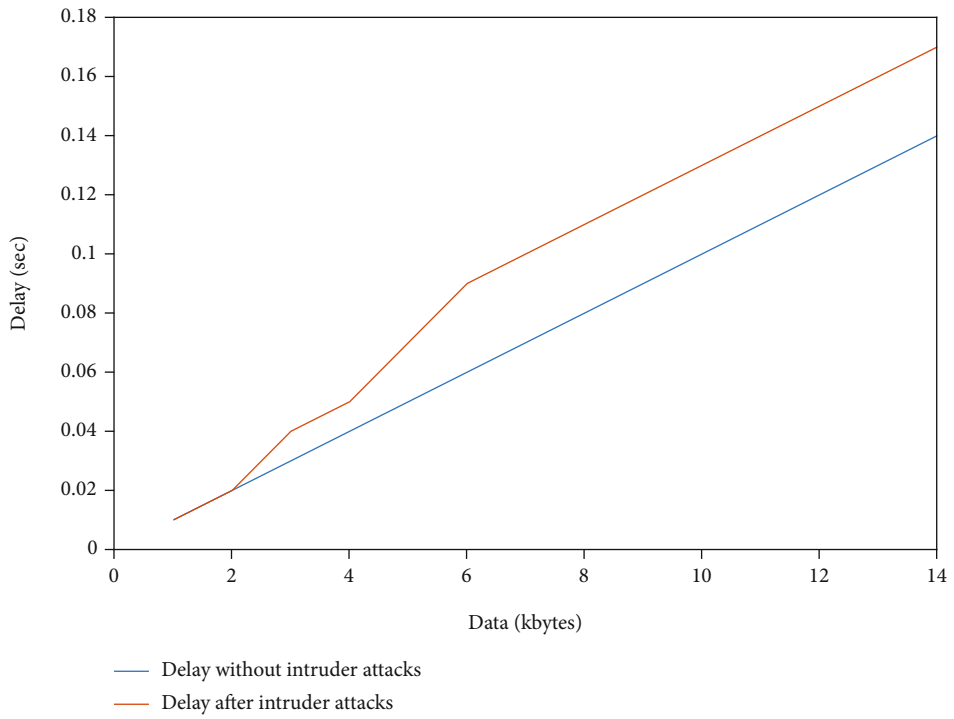


Figure 7: Transmission delay with and without intruder's attack.

Table 2. The proposed scheme detects intruder's attack by exposing collision and exhaustion probabilities as described in Section 3.

Results shown in Figure 4 describe the scenario when there is no intruder's attack found. In this case, the probability of successful transmission is maximum, and the probability of collision is 0. The sum of both probabilities is less than the threshold, and hence, no collision detection is found in this case. However, the probability of collision and successful transmission increases in the presence of an intruder's attack

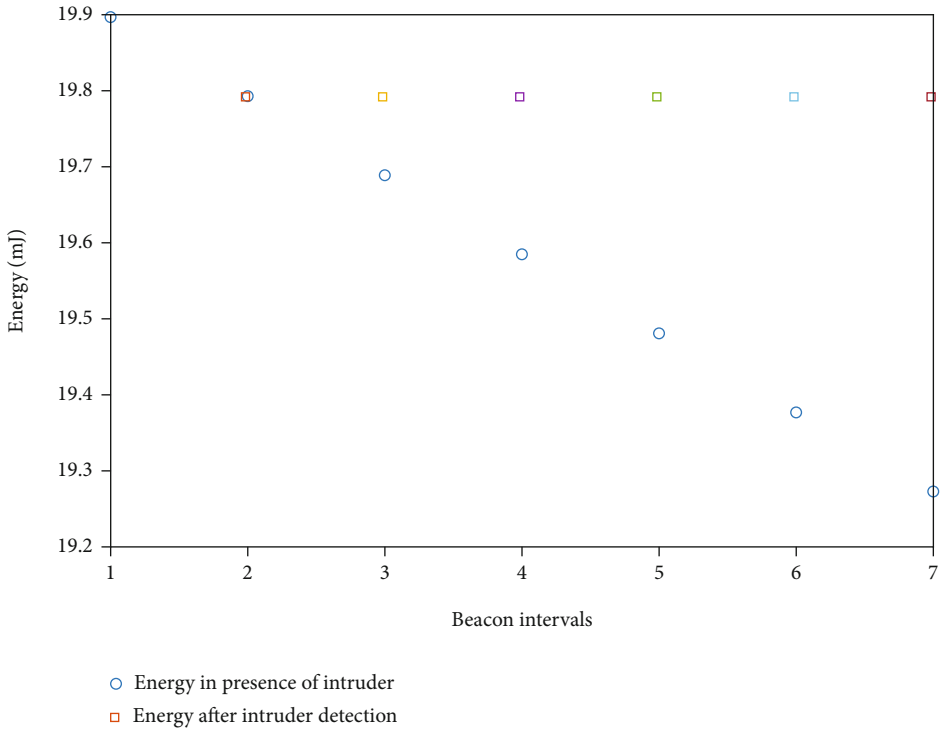


Figure 8: Energy consumed by the node for intrusion detection.

and is verified by the results shown in Figure 5. It is evident from the results that the probability of successful transmission depends upon the probability of collisions. In the 3rd and 5th beacon intervals, the probability of successful transmission decreases with the increase in collision probability.

Figure 6 shows the probabilities of exhaustion attack in the presence of intruders for different beacon intervals. It is evident from the results that whenever exhaustion probabilities increase from the threshold, our proposed scheme detects the presence of intruders.

Delay transmission is calculated as the time duration when a node has data to transmit and successfully transmits its data to the coordinator or sink. Network delay is the accumulated delay of all the nodes in a network. Figure 7 shows the amount of delay experienced by the nodes in the presence and absence of an intruder's attack. It is evident from the results that network delay in the presence of an intruder's attack is much more than the delay without an intruder's attack. This is due to the retransmission of the crushed packet.

Figure 8 shows a comparative analysis of energy consumed by a node, when it detects intruder's attacks and stops transmission, and when it could not detect an intruder and keep on transmitting packets again and fruitlessly wasting its energy.

5. Conclusion

IoT applications generate big data for remote sensing applications and thus vulnerable to adversaries, and server attacks which can severely damage the network performance and QoS may be compromised. In this work, an intrusion detection mechanism is proposed to detect intrusion possibilities by developing a soft function for the IEEE 802.15.4 standard. The proposed intrusion detection mechanism detects the collision and exhaustion attacks during different stages of the standard's superframe. Besides, if QoS is compromised due to generation of large data, then the duty cycle adaptation algorithm is proposed to meet the QoS requirements. The results verify that the proposed scheme successfully detects these intrusions, and this detection minimizes the WSN delay and energy consumption. Although the proposed intrusion detection mechanism detects the most common intruder's attacks, it does not work when adversaries attack the IoT nodes unfairly. The future work will explore efficient intrusion detection mechanisms for unfair attacks.

Data Availability

Data is available on request from the corresponding author.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: 19-ENG-1-01-0015.

References

[1] S. Khan, A. N. Alvi, M. A. Javed, B. Roh, and J. Ali, "An efficient superframe structure with optimal bandwidth utilization and reduced delay for Internet of Things based wireless sensor networks," *Sensors*, vol. 20, no. 7, p. 1971, 2020.

[2] M. Zheng, C. Wang, M. du, L. Chen, W. Liang, and H. Yu, "A short preamble cognitive MAC protocol in cognitive radio sensor networks," *IEEE Sensors Journal*, vol. 19, no. 15, pp. 6530–6538, 2019.

[3] S. Khan, A. N. Alvi, M. A. Javed, S. H. Bouk, and S. H. Bouk, "An enhanced superframe structure of IEEE 802.15.4 standard for adaptive data requirement," *Computer Communications*, vol. 169, pp. 59–70, 2021.

[4] A. N. Alvi, S. Khan, M. A. Javed et al., "OGMAD: optimal GTS-allocation mechanism for adaptive data requirements in IEEE 802.15.4 based Internet of Things," *IEEE Access*, vol. 7, pp. 170629–170639, 2019.

[5] R. Zhu, M. Yu, Y. Li, J. Wang, and L. Liu, "Edge sensing-enabled multistage hierarchical clustering deredundancy algorithm in WSNs," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6664324, 14 pages, 2021.

[6] H. Wang, L. Wu, Q. Zhao, Y. Wei, and H. Jiang, "Energy balanced source location privacy scheme using multibranch path in WSNs for IoT," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6654427, 12 pages, 2021.

[7] W. Zhang, W. Fang, Q. Zhao, X. Ji, and G. Jia, "Energy efficiency in Internet of Things: an overview," *Computers, Materials & Continua*, vol. 63, no. 2, pp. 787–811, 2020.

[8] S. Gong, X. Liu, K. Zheng, X. Tian, and Y. Zhu, "Slot-hitting ratio-based TDMA schedule for hybrid energy-harvesting wireless sensor networks," *IET Communications*, vol. 14, no. 12, pp. 1949–1956, 2020.

[9] Y. Li, X. Zhang, J. Zeng, Y. Wan, and F. Ma, "A distributed TDMA scheduling algorithm based on energy-topology factor in Internet of Things," *IEEE Access*, vol. 5, pp. 10757–10768, 2017.

[10] R. A. Lara-Cueva, R. Gordillo, L. Valencia, and D. S. Benítez, "Determining the main CSMA parameters for adequate performance of WSN for real-time volcano monitoring system applications," *IEEE Sensors Journal*, vol. 17, no. 5, pp. 1493–1502, 2017.

[11] E. B. Hamida, M. A. Javed, and W. Znaidi, "Adaptive security provisioning for vehicular safety applications," *International Journal of Space-Based and Situated Computing*, vol. 7, no. 1, pp. 16–31, 2017.

[12] M. A. Javed, E. B. Hamida, A. al-Fuqaha, and B. Bhargava, "Adaptive security for intelligent transport system applications," *IEEE Intelligent Transport Systems Magazine*, vol. 10, no. 2, pp. 110–120, 2018.

[13] A. Haider, M. Adnan Khan, A. Rehman, M. U. Rahman, and H. Seok Kim, "A real-time sequential deep extreme learning machine cybersecurity intrusion detection system," *Computers, Materials & Continua*, vol. 66, no. 2, pp. 1785–1798, 2021.

[14] J. Yu, E. Kim, H. Kim, and J. H. Huh, "Design of a framework to detect device spoofing attacks using network characteristics," *IEEE Consumer Electronics Magazine*, vol. 9, no. 2, pp. 34–40, 2020.

[15] E. Yang, G. Prasad Joshi, and C. Seo, "Improving the detection rate of rarely appearing intrusions in network-based intrusion detection systems," *Computers, Materials & Continua*, vol. 66, no. 2, pp. 1647–1663, 2021.

[16] M. Eskandari, Z. H. Janjua, M. Vecchio, and F. Antonelli, "Passban IDS: an intelligent anomaly-based intrusion detection system for IoT edge devices," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6882–6897, 2020.

[17] E. Anthi, L. Williams, M. Słowińska, G. Theodorakopoulos, and P. Burnap, "A supervised intrusion detection system for smart home IoT devices," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9042–9053, 2019.

[18] M. Mittal and S. Vijayal, "Detection of attacks in IoT based on ontology using SPARQL," in *2017 7th International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 206–211, Nagpur, India, 2017.

[19] M. Mittal, L. K. Saraswat, C. Iwendi, and J. H. Anajemba, "A neuro-fuzzy approach for intrusion detection in energy efficient sensor routing," in *2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, pp. 1–5, Ghaziabad, India, 2019.

[20] M. Mittal, C. Iwendi, S. Khan, and R. Javed, "Analysis of security and energy efficiency for shortest route discovery in low-energy adaptive clustering hierarchy protocol using Levenberg-Marquardt neural network and gated recurrent unit for intrusion detection system," in *Transactions on Emerging Telecommunications Technologies*, John Wiley & Sons, Ltd., 2020.

[21] U. M. Malik, M. A. Javed, S. Zeadally, and S. U. Islam, "Energy efficient fog computing for 6G enabled massive IoT: recent trends and future opportunities," *IEEE Internet of Things Journal*, 2021.

[22] N. U. H. Lodhi, A. Malik, T. Zulfiqar, M. A. Javed, and N. S. Nafi, "Performance evaluation of Wi-Fi finger printing based indoor positioning system," in *2018 IEEE Conference on Wireless Sensors (ICWiSe)*, pp. 56–61, Langkawi, Malaysia, 2018.

[23] M. A. Javed and S. Zeadally, "RepGuide: reputation-based route guidance using Internet of Vehicles," *IEEE Communications Standards Magazine*, vol. 2, no. 4, pp. 81–87, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 1039613, 11 pages
https://doi.org/10.1155/2021/1039613

Research Article

Design and Implementation of a Robust Convolutional Neural Network-Based Traffic Matrix Estimator for Cloud Networks

Rashida Ali Memon[1], **Sameer Qazi**[2] and **Bilal Muhammad Khan**[1]

[1]*Department of Electronics & Power Engineering, PN Engineering College, National University of Sciences & Technology, Habib Ibrahim Road, Karachi 75350, Pakistan*
[2]*Department of Electrical Engineering, College of Engineering, PAF Karachi Institute of Economics & Technology (PAF-KIET), Korangi Creek, Karachi 75190, Pakistan*

Correspondence should be addressed to Rashida Ali Memon; rashida@pnec.nust.edu.pk

Received 16 April 2021; Accepted 10 May 2021; Published 3 June 2021

Academic Editor: Nawab Muhammad Faseeh Qureshi

Recent research literature shows promising results by convolutional neural network- (CNN-) based approaches for estimation of traffic matrix of cloud networks using different architectures. Although conventionally, convolutional neural network-based approaches yield superior estimation; however, these rely on assumptions of availability of a large training dataset which is completely accurate and nonsparse. In real world, both these assumptions are problematic as training data size may be limited, and it is also prone to missing (or incomplete) measurements as well as may have measurement errors. Similarly, the 2-D training datasets derived from network topology based may be sparse. We investigate these challenges and develop a novel architecture which can cater for these challenges and deliver superior performance. Our approach shows promising results for traffic matrix estimation using convolutional neural network-based techniques in the presence of limited training data and outlier measurements.

1. Introduction

Internet traffic load is increasing multifold every few years. In US alone, only the business internet traffic volume has seen a rapidly increasing trend growing from 45.1 billion GB in 2016 to 112.7 billion GB in 2020 and projected to increase up to 224.8 billion GB in 2023 [1]. Two technologies are primary contributors for this increased traffic load: (1) cloud computing services which conveniently offer the Internet as a one-stop solution to all users in the form of infrastructure, platform, and software as service (AaaS, PaaS, and SaaS). (2) software-defined networks (SDNs) devise intelligent technologies to cater for ensuring Quality of Service (QoS) for Internet users even in presence of an ever-increasing Internet cross traffic. This places significant burden on the network as traditional traffic engineering (TE) techniques are effectively bypassed in software-defined networks. Akyildis et al. provide a comprehensive review of a roadmap [2] and challenges [3] to implement TE in SDNs in their survey papers. While the traditional flow level-based routing strategies to ensure QoS like MPLS, ATM, IntServ, and DiffServ failed in achieving scalability and QoS guarantees, SDNs employ programmable OpenFlow (OF) switches that communicate with SDN controllers via OF protocol to dynamically modify the traditional forwarding tables of routers based on a flow-level control in scalable manner such as using hash-based approaches and employing more powerful hardware like multithreading controllers. The absence of an efficient network monitoring system (NMS) also makes the network vulnerable to security threats like DoS/DDoS [4] and Sybil attacks [5]. An NMS can timely detect and thwart such threats by timely detection of network traffic anomalies [6].

With conventional traffic engineering (TE) solutions bypassed in SDNs, it is only a matter of time when SDNs would have squeezed maximum QoS out of the network necessitating further capacity planning by network administrators. Thus, it becomes increasingly important to estimate the network parameters for an improved and reliable

network capacity planning and management. Estimation of dynamic live network parameters has been investigated by a lot of researchers [7]. Special emphasis on parameters affecting the QoS of network includes network delays [8], network traffic volumes [9], packet losses, and network capacity planning. Network delays are important as some network applications like voice over IP (VoIP) require end-to-end delays to be within a specified threshold usually, and the limit for one-way delay of data packets carrying voice should be less than 150 ms. Similarly, VoIP requires packet loss rates of less than 1% and average one-way packet jitter to be less than 30 msec. Similarly, capacity planning in the form of upgrading network bandwidth and/or employment of effective traffic shaping/policing techniques is equally important as in the absence of that rate adaptive flows may be hurt by nonrate adaptive flows [10].

Recent research literature shows promising results by using neural network-based approaches for estimation of traffic matrix using different architectures, such as use of Artificial Neural Networks (ANNs) [11] and recurrent neural networks (RNNs) [12]. Another new approach used recently is convolutional neural network- (CNN-) based which shows promising results for traffic matrix estimation. Although the results of traffic matrix prediction using a CNN-based architecture are impressive under ideal assumed conditions, the results will deteriorate if real-world problems are considered. We list ideal conditions here: (1) an assumption that a huge volume of training dataset is available, and this is often not the case in real situations; (2) an assumption that training dataset is complete, i.e., it is not having any missing measurements, measurement noise, or outliers.

A consequence of the violation of any of the above assumptions is that if we employ CNN using standard architectures, they may yield suboptimal performance, e.g., if a small volume of dataset is available for training or if we use an adaptive learning-rate optimizer such as Adagrad as used in [13], and it will yield inferior performance because of the incorrect assumption of the availability of large clean dataset; it will adaptively lower the learning rate as the training would progress over time.

In this paper, we develop a robust CNN-based traffic matrix estimation framework which can deliver superior performance in the presence of limiting training data or training data with outlier measurements.

The rest of this paper is organized as follows. Section II discusses the related work. We present the problem of traffic matrix estimation in Section III. Section IV outlines the methodology of our techniques, followed by Section V in which we present the simulation results, and finally, Section VI draws the conclusions.

2. Related Work

Network parameter measurement techniques have been covered exhaustively in research literature. Few popular technique-specific terms coined by researchers include Kriging [14], cartography [15], tomography [16], and compressed sensing [17, 18]. Technically, all of these network parameter estimation techniques can be separated as those which are space-based, time-based, and use spatiotemporal techniques [19]. The temporal aspect of the measured network parameters is obtained through the availability of network training data while the spatial aspect is generated from the knowledge of topology of the network and how it can potentially impact the measured parameters. The network parameter estimation or prediction algorithm in these spatiotemporal techniques belongs to approaches such as Bayesian estimation, Linear Optimization, and Maximum Likelihood (or Expectation Maximization) [9].

It is pertinent to highlight that for the problem under consideration in this paper, namely, traffic matrix estimation, the accuracy of training data is of utmost importance. For example, the expectation maximum-based estimation approach uses statistical inference tools that are based on distribution of elements of the traffic matrix. These are then used to compute, based on information of link counts, an expectation of the elements of the traffic matrix. Such techniques, however, rely heavily on initial or given knowledge of the traffic flow mean and variance. Many researchers, for this purpose, have utilized various distributions such as Gaussian and Poisson. [20]. Towards that end, even in the case when initial or preliminary estimates of a recognized distribution of traffic flows are known, these techniques very much rely on knowledge of the initial prior [9], which acts as a starting point for optimization algorithms to yield a solution obeying all the spatial constraints of the problem. Hence, these techniques are highly dependent upon 'goodness' of a known or prior solution. Earlier research work done by Tebaldi et al. [21] and Vardi [22] for estimation of traffic elements in IP backbones introduced the idea of adapting a hybrid approach, in which elements of the traffic matrix are either estimated or algebraically computed based on link load measurements. This approach significantly reduces complexity of the problem; since in this case, estimation of a much lesser number of origin/destination (OD) pair traffic flows is required. Such schemes have shown to be beneficial in cases where the routing matrix exhibits a highly decaying Eigen spectrum [23]. A few noteworthy research contributions related to problem of traffic matrix estimation include Bayesian Learning techniques which have been carried out by Nie [6] and Fan [24].

Recent research literature shows promising results by neural network-based approaches for estimate of traffic matrix using different architectures. Jiang et al. [11] fuse the neural network approach with time frequency analysis for network traffic matrix estimation. Emami et al. [13] have devised a convolutional neural network- (CNN-) based traffic matrix estimation architecture in which one part of the training data (link loads) is transformed into one higher dimension by considering the network topology. This enables exploitation of convolutional neural network-based techniques which are optimized for image (2D) datasets and are considered more robust than other types of neural networks.

Qian et al. [12] present another technique for traffic matrix estimation without any training data using only current but partial or incomplete OD flow data. The technique exploits recurrent neural networks (RNNs) to estimate the unknown OD flows from known OD flow measurements.

Qazi et al. [18] have developed a compressed sensing model for network traffic matrix estimation based on a dynamic network measurement model. The model is based on traffic demands instead of a stationary routing matrix. This technique shows that link count measurements could be further reduced with traffic matrix estimation with tolerable errors.

3. The Traffic Matrix Estimation Problem

The traffic matrix estimation problem can be represented by the following relationship:

$$Y = AX \tag{1}$$

where $Y \in \Re^{mxt}$ is the observable matrix of measured link counts, and m is the number of network links, recorded at indexed time intervals. $A \in \Re^{m \times n^2}$ is called routing matrix that expresses routing configurations where m and n denote the number of wide area links and those of cloud nodes, respectively. These result in n^2 possible paths between cloud nodes. We assume that the routing matrix is stationary, i.e., remains stable (unlike typical wireless networks) over long periods of time. This is valid especially for the case of Internet; since in this case, majority of the wide area paths can remain stable for several hours or days. Finally, $X \in \Re^{n\wedge 2xt}$ is an unobservable traffic matrix which is required to be estimated over the specified time intervals as indexed by t.

Solution to the traffic matrix problem as given in Equation (1) is oftentimes quite a challenging one. This is because of a variety of reasons: (a) the routing matrix A is a "fat" matrix; and since $m << n$, the system of linear equations under consideration is ill-posed, underdetermined, and underconstrained. It is pertinent to highlight the fact that for an underdetermined system of equations, a unique solution may not exist. On the other hand, if we have a balanced, well-posed, or an overdetermined linear system of equations (e.g., if we have A as a 'tall' or a 'square' matrix, i.e., when $m \geq n$), then we may have a unique solution. (b) In this modern age of dynamic software defined, validity of assumption for A being stationary is no longer acceptable. Furthermore, in these modern times, routing is no longer dictated by any deterministic algorithms. Instead, more "opportunistic" software-defined strategies are being used to allow for user-centric routing decisions rather than being network-centric in nature. In case of software-defined cloud computing platforms, several decisions while being user-centric are still treated as network oriented in order to obtain good load balancing on the network servers or the network as a whole. 1st generation research works using the assumption of stationary routing matrix A have investigated range of spatiotemporal methods as explained in the previous section. More recent works, e.g., [13] have devised strategies in which the topology information is embedded into the model using flexibility, e.g., by using the link load measurements and embedding them into link adjacency matrix. This relaxes the spatial constraints in the original problem while still enabling accurate estimation through an efficient learning process of a neural network.

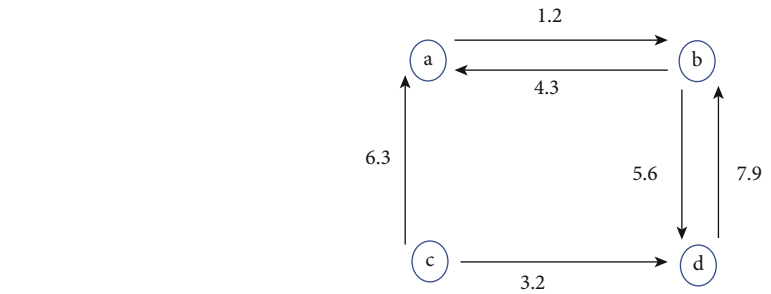
It is a well-known fact that in network traffic estimation problem, quick anomaly detection is one of the prime motivators of the development of these systems; hence, the estimator should have superior temporal prediction performance.

4. Robust CNN-Based Traffic Matrix Estimation

Recent work by Emami et al. [13] incorporating graph embedding and convolutional neural network proposed an idea to fuse the topology of network into the neural network-based estimation framework by using the graph embedding approach. This enables to add one dimension to the training data. The training data is traditionally composed of two, 1D vectors of link loads and OD flows. Using the above mentioned graph embedding technique, the observable data of link loads is transformed from 1D measurements into 2D measurements by implicit introduction of the topology into the measurement framework. This is through a 2D link adjacency matrix referred to as L2AM [13]. Figure 1 shows this graph embedding technique using a toy example as well as the CNN-based architecture to learn OD flow features using this L2AM matrix training dataset. In the original Emami et al. [13] architecture as shown in Figure 1, the output of the convolution part consists of N matrices which are vectorized to generate N different feature vectors of size $W \times W$ (corresponding to each OD flow) fed to N parallel fully connected networks (FCN) with one hidden layer and one output layer corresponding to each OD flow. The values of L and W are 13 and 9, respectively. The architecture that proposed (CNTME) gives superior performance over other contemporary approaches; however, we find that the performance becomes suboptimal when we consider that 2D training data supplied via the network topology information through link adjacency matrix is sparse (since number of OD pairs is much greater than the number of links in the network), and further, it may also have errors or noise in it. This can cause the CNN-based estimator to have suboptimal performance in terms of feature learning about the OD flow or generate outlier predictions especially when the problem at hand is a regression problem and not a classification problem. Motivated by the findings above, we see that the CNN-based traffic matrix estimator is not robust to outliers or errors in the training data.

We propose to rewire the original architecture so that the final feature set learned is that of the entire system rather than of individual OD flows. This has the effect of diluting any learning errors causing incorrect features to be learnt about individual OD flows. We call this new proposed architecture as R-CNTME. We present more details about this in the following paragraph.

If the 2D training data has limited measurements or outliers, CNTME architecture has the effect of magnifying the errors in the estimation of each individual OD flows via N parallel FCN. Our architecture differs that the feature matrices generated after the downsampled output of the convolution layers are flattened before the FCN layer (instead of afterwards) to generate one feature vector of the entire system (N OD flows) instead of individual OD flows. This has



Src\Dst	a	b	c	d
a	0	1.2	0	0
b	4.3	0	0	5.6
c	6.3	0	0	3.2
d	0	7.9	0	0

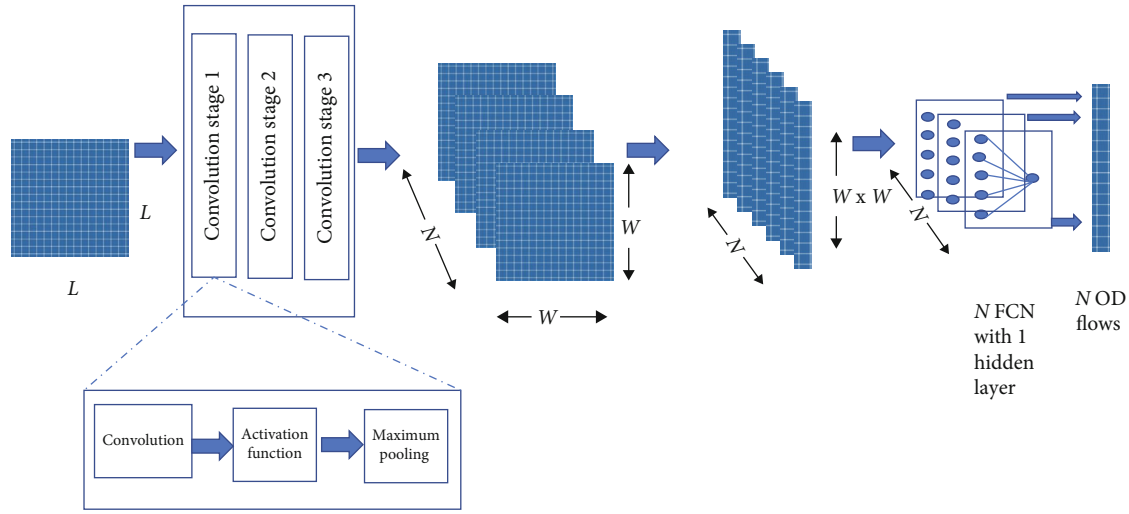


Figure 1: (top) Graph embedding technique using link adjacency matrix and numbers over arrows shows traffic volume over directed links; (bottom) convolutional neural network technique (CNTME) proposed by Emami et al. [13].

the effect of diluting the learning errors in the CNN stage caused due to sparsity of 2D training data, limited training data, and potential outlier measurements in it. Detailed schematic of the proposed architecture is given in Figure 2 below.

5. Description of ABILENE Dataset

We use the real dataset of the Abilene network (a backbone network in the United States) to evaluate the performance of our proposed ROBUST CNN-based technique. This network consists of a total of 12 nodes. Accordingly, there are a total of 144 OD flows (since $n = 12 \Rightarrow N = 12^2 = 144$). There are a total of 54 links (i.e., $m = 54$), of which 30 links provide for the connectivity between the near-neighbor nodes. The rest connects all other nodes that are available over the Internet. The Abilene network dataset consists of end-to-end traffic measurement for 24 weeks. Since the measurement time slots are considered as 5-minute intervals, therefore, there are a total of 2016 measurement points (i.e., 7 times 24 times 12 which equals to 2016) per week.

In this work, we consider the external network as an independent node. We add it with the 12 nodes of the Abilene network to cater for the load between the Abilene network and the external network. Accordingly, the L2AMs become matrices of 13×13 size.

5.1. Performance Evaluation. To test the performance of the traffic matrix estimation architecture proposed in this paper and to do a fair comparison with Emami et al. [13], we use Adam optimizer rather than Adagrad because we also simulate noise in the training data, and Adagrad lowers its learning rates with training epochs. We use mean squared error (MSE) as the loss function.

Like Emami et al. [13], we use the Softplus activation functions for the benefit that it does not give the problem of negative estimated values; so, no additional steps are needed for negative values treatment. Softplus generates output values as per function $f(x)$ for input values x.

$$f(x) = \log\left(1 + e^{x}\right) \tag{2}$$

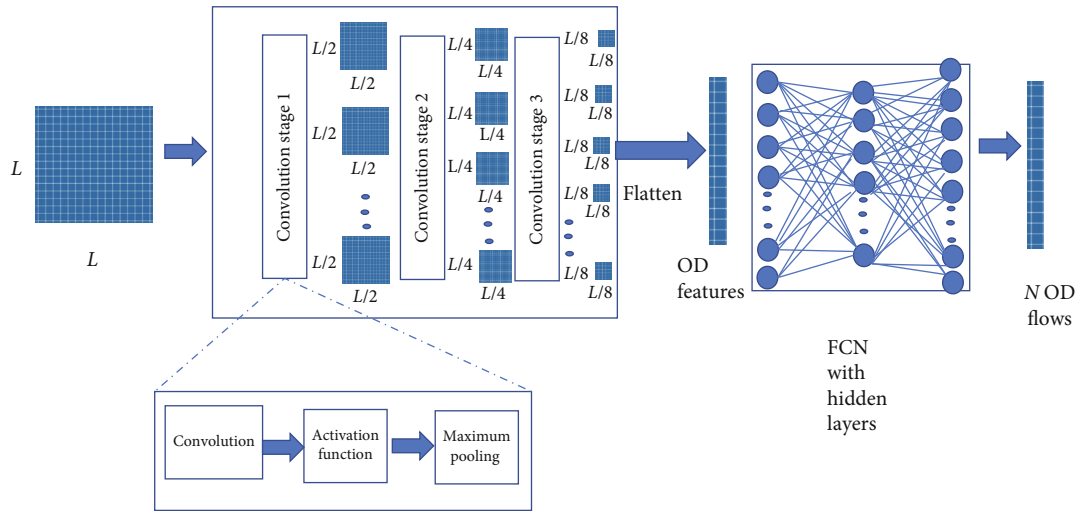


Figure 2: Schematic of the proposed architecture for robust convolutional neural network-based traffic matrix estimator (R-CNTME).

Performance of the proposed techniques has been evaluated using Abilene network datasets based on real captured network traffic as described in the previous section. For this purpose, we have used two different platforms, namely, jupyter notebook (online) and spyder python 3.7 at AMD Radeon R7 240 with a PC. The specifications of the PC are as follows. CPU: Intel Core i7 7770 @3.6 GHz with 32 GB RAM (DDR4). We have extensively evaluated all our simulations on Colab, which uses GPU Tesla K80 for TensorFlow and Keras. In all the considered cases, we use the first 1500 samples of the first week data as training dataset and the last 500 samples for testing unless specified explicitly. To compare the performance of our scheme, we use the established metrics that are widely used in the relevant recent research literature [25]. These metrics are defined as follows:

Spatial relative error (SRE):

$$\mathrm{SRE}(i) = \frac{\left\|\widehat{X}_i - X_i\right\|_2}{\|X_i\|_2}; i = 1, 2, \cdots, n^2 \tag{3}$$

where X_i represents the actual OD flows, and $\widehat{X}_i$ represents estimated OD flows for $i \in 1, 2, \cdots, n^2$ at time t.

Temporal relative error (TRE):

$$\mathrm{TRE}(t) = \frac{\left\|\widehat{X}_t - X_t\right\|_2}{\|X_t\|_2}; t = 1, 2, \cdots, T \tag{4}$$

in which X_t represents the actual OD flow,and $\widehat{X}_t$ represents estimated OD flow for time t while T denotes the length of the period of testing.

Bias:

$$\mathrm{Bias}(i) = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{X}_{i,t} - X_{i,t}\right) \tag{5}$$

Table 1: Simulated errors in the L2AM training data.

Error scenario	% of Noisy L2AM entries	Noise locations in all time indices	Noise distribution
E1	30%	Random	Gaussian N(0, 25e12)

Table 2: Number of CNN and FCN layers.

CNN layers (filter size 2×2)	FCN hidden layers
3	81 : 144
2	81 : 100 : 144
2	81 : 100 : 121 : 144
2	81 : 90 : 100:121 : 144
1	81 : 144

Standard deviation:

$$\text{Standard Deviation}\,(i) = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(\left(\widehat{X}_{i,t} - X_{i,t}\right) - \mathrm{bias}(i)\right)^2} \tag{6}$$

where $X_{i,t}$ represents the actual OD flow, and $\widehat{X}_{i,t}$ represents the estimated OD flow for $i \in 1, 2, \cdots, n^2$at time t, while T denotes the length of the period of testing.

We find that this framework works well when there are no errors in the training data. We generate random noise for the L2AM matrix training data to 30% of the nonzero entries, respectively. We call it error scenario (E1) in which we generate noise in random 30% locations of the nonzero entries of the training data using the Gaussian distribution with zero mean and standard deviation equal to 5×10^6.

5.2. Performance of Proposed Architecture with Different Numbers of CNN and FCN Layers with Simulated Errors in Abilene Dataset. To validate our architecture, we test the

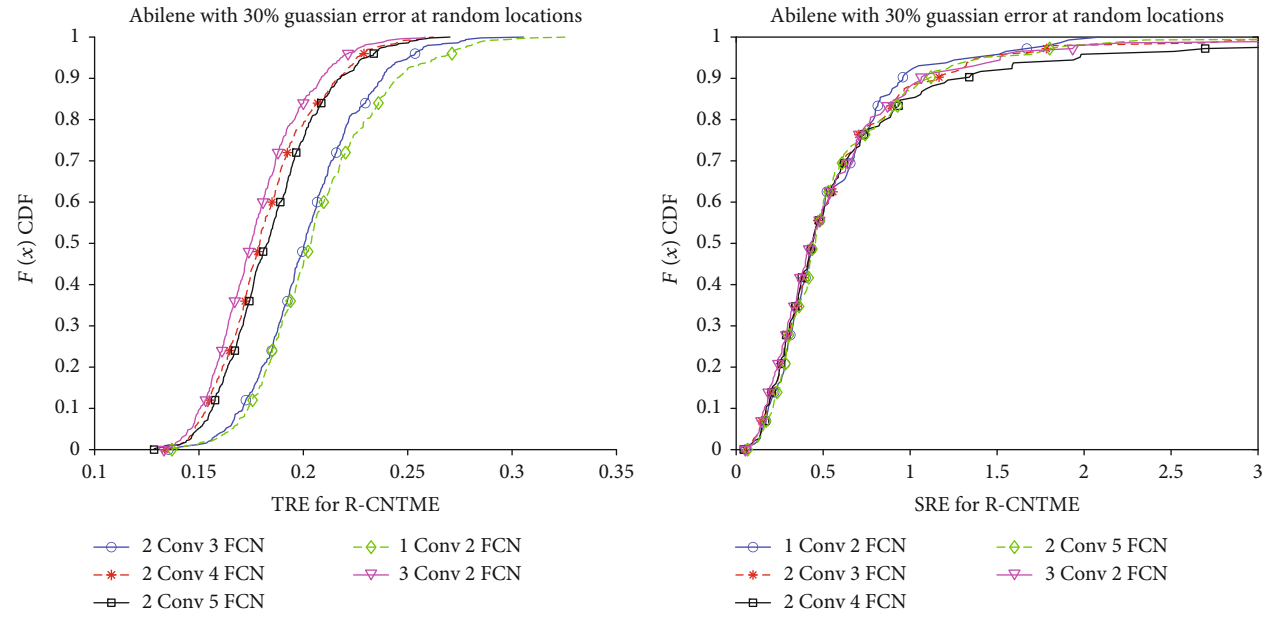


Figure 3: SRE and TRE performance of proposed R-CNTME architecture when errors are introduced in the training data and number of CNN layers and FCN layers are varied.

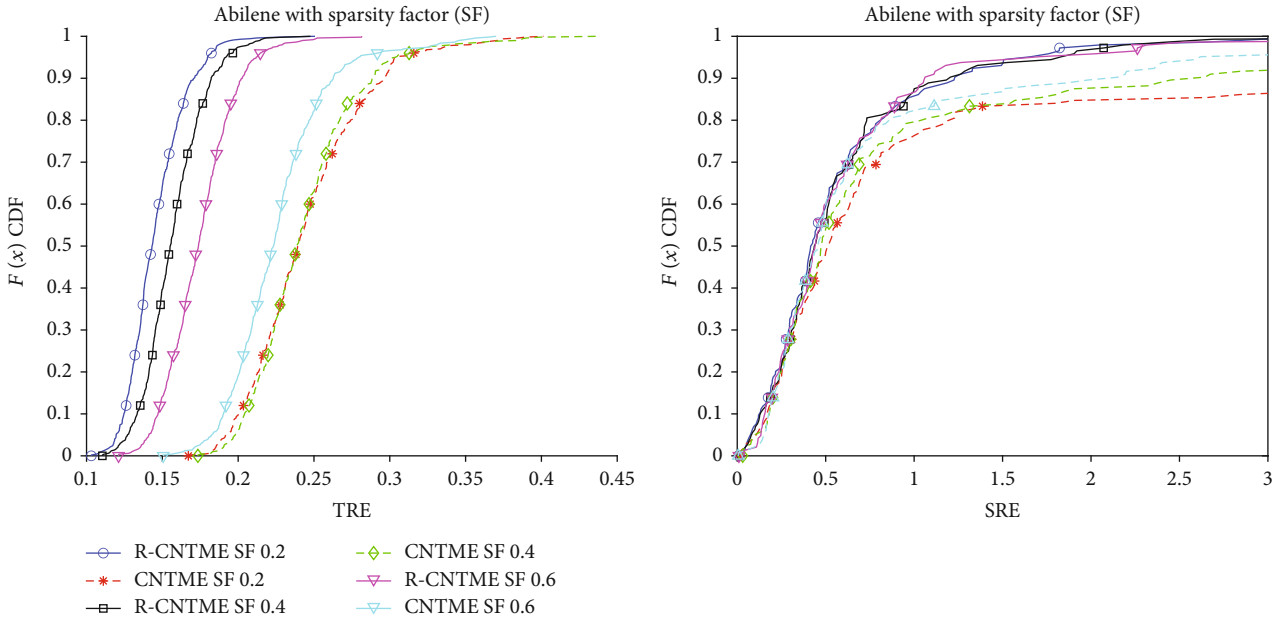


Figure 4: Comparison of performance of R-CNTME AND CNTME for Abilene dataset for different sparsity.

performance of our proposed scheme by first introducing errors in the 2D training L2AM data using simulated error scenario E1 as shown in Table 1. This is so that we can simulate both the effects of sparsity and training data errors while validating the model. We vary the number of CNN and FCN layers from 3 to 5 using the specified hidden layers as shown in Table 2.

We observe from the result in Figure 3 that as number of CNN layers and FCN layers are varied, we see marginal differences in the SRE performance. However, we notice substantial improvement in TRE performance (as CNN and FCN layers are increased); optimum trade-off for TRE metrics is best when the number of CNN layers is 3 and number of FCN layers are equal 2. This is because increasing the number of neurons in the hidden layers of the FCN is well known to cause the problem of overfitting, and too few neurons in hidden layers are well known to cause the problem of underfitting. Hence, CNN and FCN layers cannot be increased indefinitely as network anomaly prediction requires the estimator to be good in both the spatial and temporal aspects.

Addition of more hidden layers in the FCN part may add more nonlinearities and may improve the SRE performance

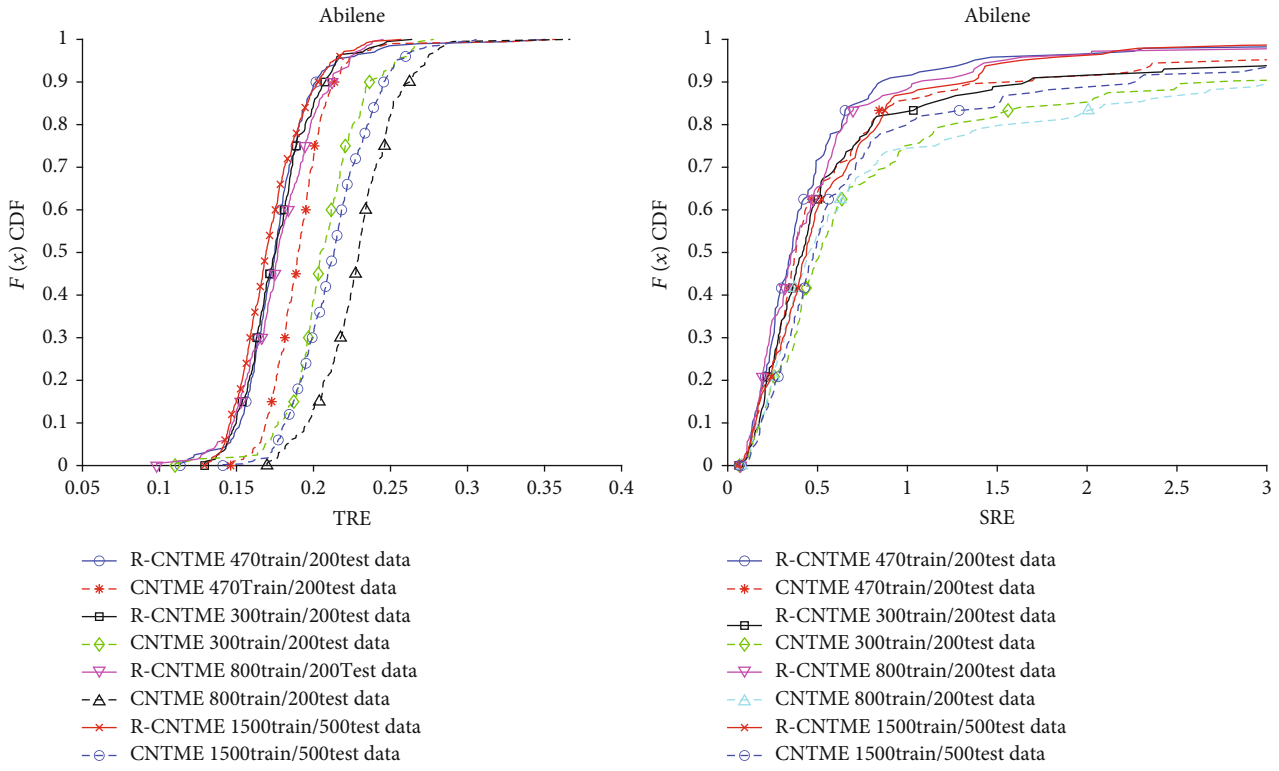


Figure 5: Comparison of performance of R-CNTME AND CNTME for the Abilene dataset for different train-test splits.

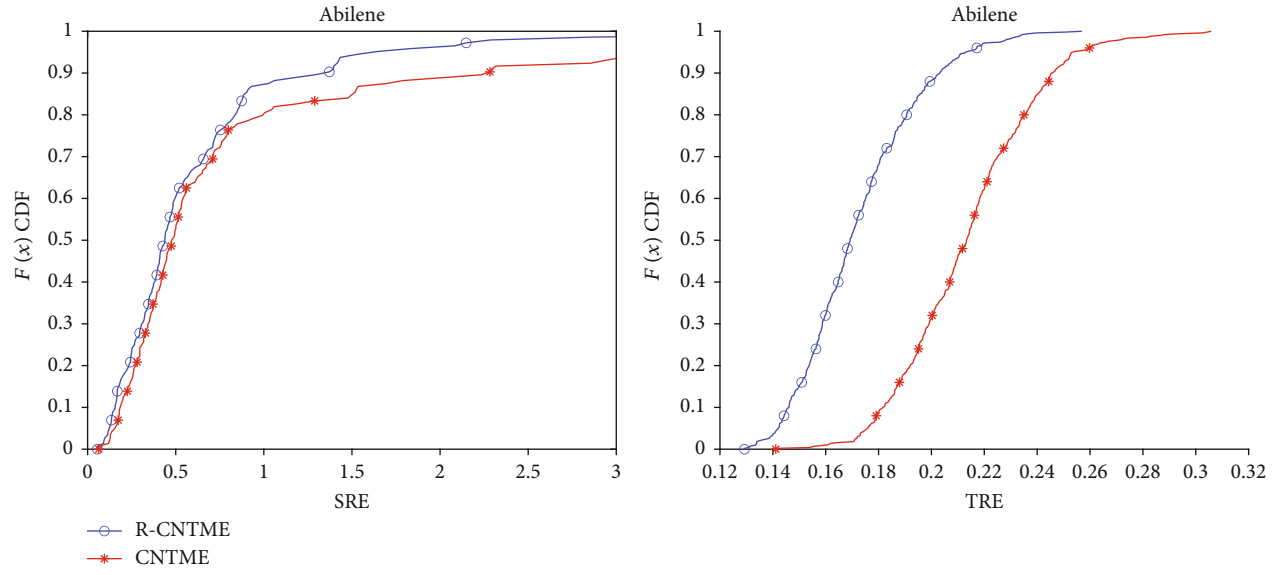


Figure 6: CDF of SRE and TRE for R-CNTME and CNTME for the Abilene dataset.

of the system at the cost of TRE performance since estimator performance becomes poorer on dataset outside of training data due to overfitting. It is pertinent to highlight here the fact that in the network traffic estimation problem, quick anomaly detection is one of the prime motivators of the development of these systems; hence, the estimator should have superior temporal prediction performance.

5.3. Comparison of Performance of R-CNTME and CNTME

5.3.1. Impact of Sparsity of Training Data. For this experiment, we generate different sparsity factors on the Abilene dataset. We synthetically modify the topology; so, it has 12 nodes, but the number of links is determined as per the sparsity factor. Sparsity factor can be defined as the number of

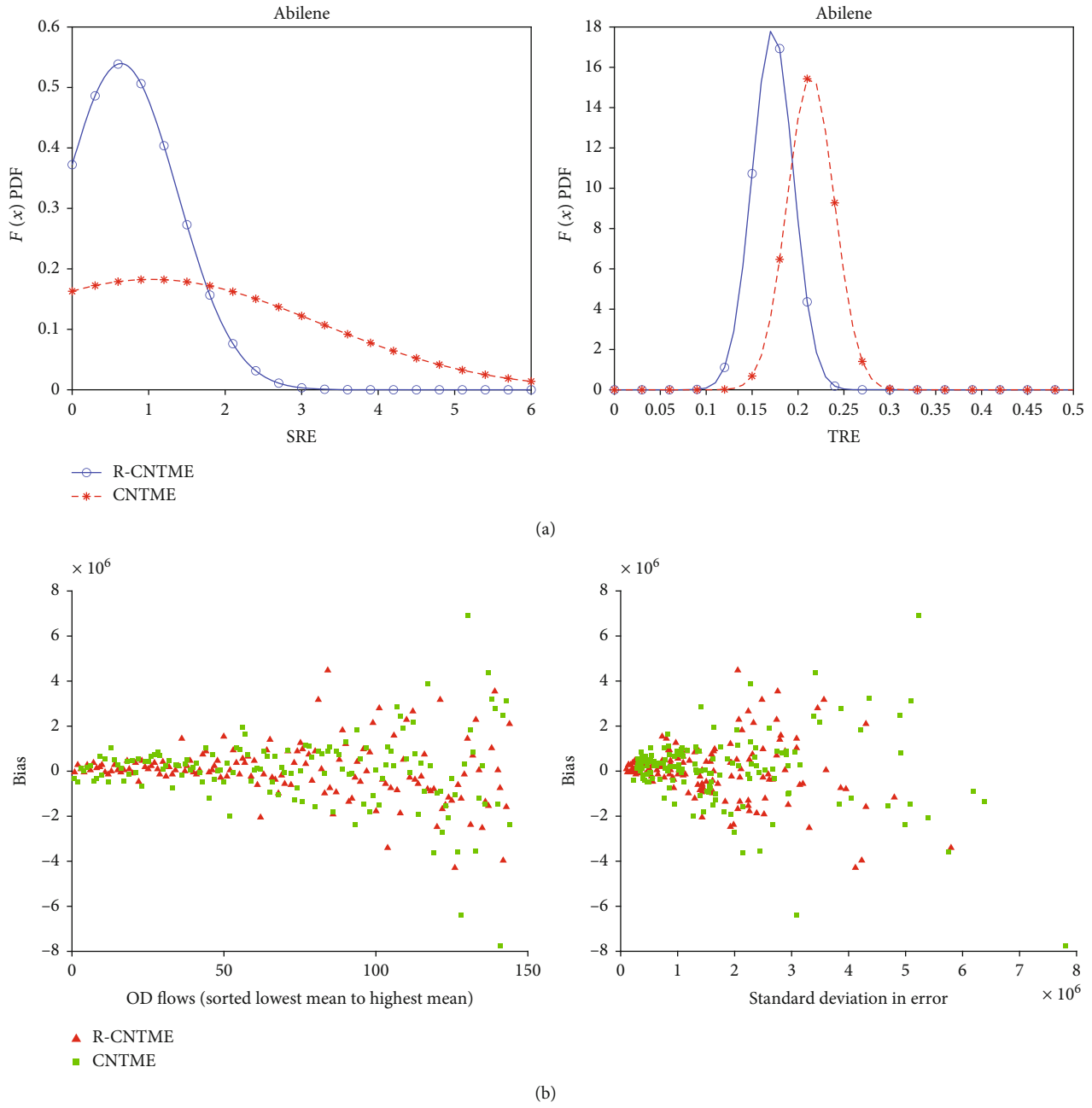


Figure 7: (a) PDF of SRE and TRE for R-CNTME and CNTME for the Abilene dataset. (b) Bias and standard deviation performance of R-CNTME and CNTME for the Abilene dataset.

zero entries to total number of entries in the L2AM training data. We ensure that the topology is sufficiently connected so that there is a path between each OD pair. Note that the OD traffic volumes between the 144 OD pairs are same as that of the original Abilene network; however, link counts are varied as per the simulated topology with variable sparsity factors. Figure 4 shows the results of this experiment. As the sparsity factor is reduced to 0.2, R-CNTME has superior TRE of 0.187 or lower for 90% of cases; for CNTME, the corresponding value is 0.27. On increasing sparsity factor to 0.6, the previous values increase to 0.2 and 0.29, respectively. SRE graphs show negligible degradation for R-CNTME and CNTME as sparsity is increased.

5.3.2. Impact of Size of Training Data. Figure 5 shows the SRE and TRE performance as the size of the training and test data is varied. It is observed from TRE results that as size of training data is increased, the TRE performance of R-CNTME is better. When the training data size is 1500, 90% of TRE values for R-CNTME are 0.19 or lower; when training data is reduced to 300, this increases to just 0.2. For CNTME, the TRE performance is lower at all ranges of training data

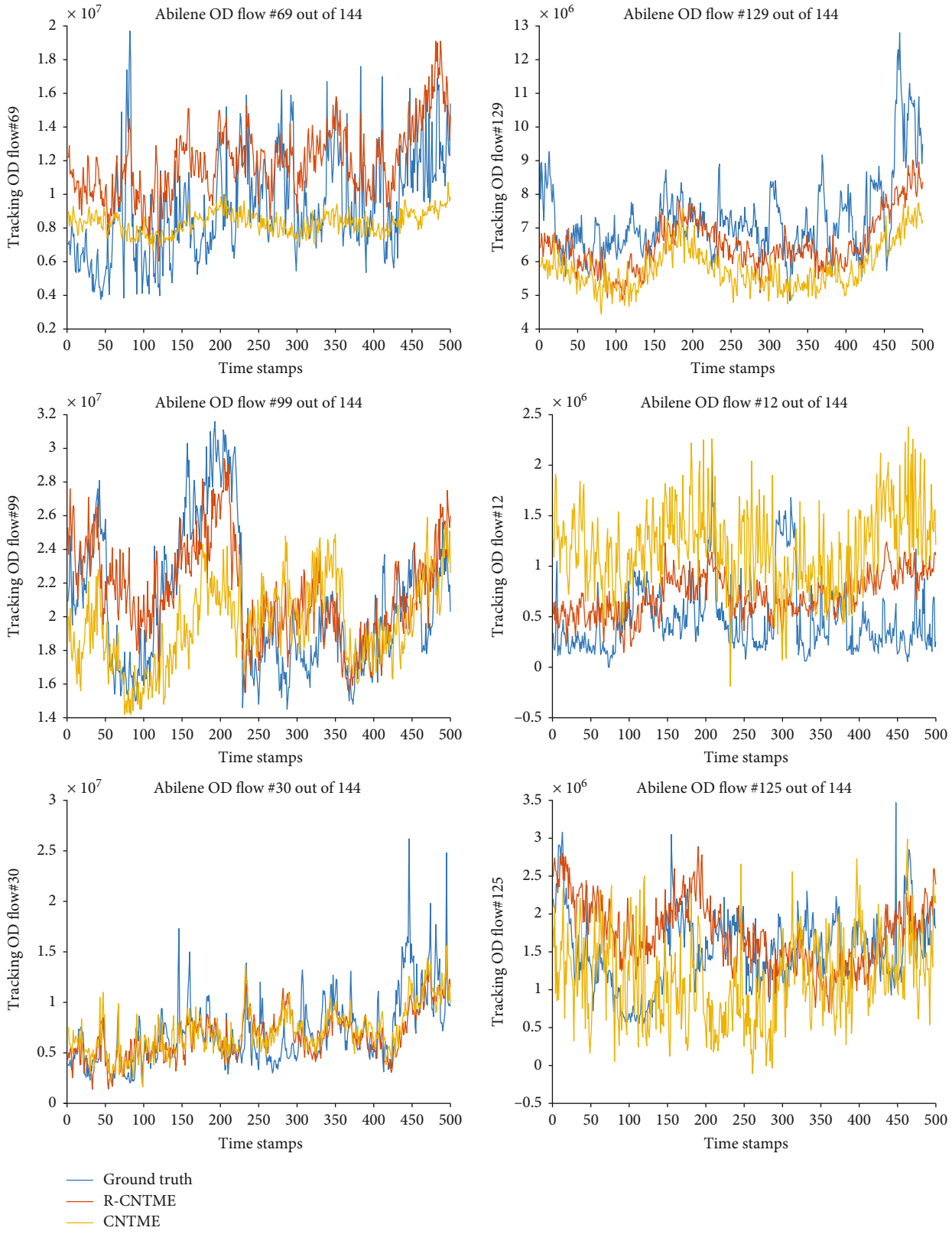


Figure 8: OD tracking performance.

sizes, and it also follows an erratic behavior due to the model being affected by underfitting and overfitting problem explained earlier. This shows the robustness of R-CNTME for limited training data availability too. The SRE performance of R-CNTME is also better than CNTME for all ranges of the training data size.

5.4. Performance Comparison of R-CNTME and CNTME on Clean Abilene Dataset. Figure 6 shows the CDF of the SRE and TRE for the clean Abilene dataset without any artificial effects of simulated errors or sparsity in the L2AM training data.

We see that the performance of R-CNTME is best for both TRE and SRE metrics. R-CNTME has a SRE value 1.5 or below for 95% of cases compared to 82% for CNTME. Similarly, the TRE performance of R-CNTME shows TRE values of less than 0.2 for 90% of the cases. The corresponding values for CNTME are 0.24. The SRE and TRE behavior is also shown as fitted to a normal pdf (Figure 7(a)). These graphs confirm the superior behavior of R-CNTME over CNTME. In addition, RCNTME is also displaying better bias and standard deviation performance than CNTME (Figure 7(b)).

5.5. OD Tracking Performance of R-CNTME and CNTME. Figure 8 shows the OD prediction performance when compared with the ground truth for six random OD flows. It is clear that R-CNTME displays superior behavior in OD flow prediction from link flow data compared to CNTME. It not only predicts the OD flows closely but also accurately tracks the anomalies in correspondence with the actual ground truth.

6. Conclusion

Accurate traffic estimation is very necessary in back-haul cloud networks for quick detection and prevention of an anomaly. In this paper, motivated by recent work for convolutional neural network-based network traffic matrix estimation, an architecture for a robust convolutional neural network-based traffic matrix estimator was proposed for Cloud Network Traffic Estimation which displays better performance in presence of sparse 2D training datasets and 2D datasets having errors and limited training dataset availability. The comprehensive simulations with real-world datasets reveal that the proposed architecture has stable performance with the training data artifacts and also exhibits superior error performance and anomaly detection performance.

Data Availability

We use the real dataset of the Abilene network (a backbone network in the United States) to evaluate the performance of our proposed technique.

Conflicts of Interest

The author(s) declare(s) that they have no conflicts of interest.

References

[1] S. O'Dea, "The economic value of Wi-Fi: A global view (2018 and 2023)," *Telecom Advisory Services; Cisco Systems*, p. 39, 2018, https://www.statista.com/statistics/995060/business-internet-traffic-in-the-us/.

[2] I. F. Akyildiz, A. Lee, P. Wang, M. Luo, and W. Chou, "A roadmap for traffic engineering in SDN-OpenFlow networks," *Computer Networks*, vol. 71, pp. 1–30, 2014.

[3] I. F. Akyildiz, A. Lee, P. Wang, M. Luo, and W. Chou, "Research challenges for traffic engineering in software defined networks," *IEEE Network*, vol. 30, no. 3, pp. 52–58, 2016.

[4] M. A. Saleh and A. Abdul Manaf, "Optimal specifications for a protective framework against HTTP-based DoS and DDoS attacks," in *2014 International Symposium on Biometrics and Security Technologies (ISBAST)*, pp. 263–267, Kuala Lumpur, Malaysia, Aug. 2014.

[5] J. Mao, X. Li, Q. Lin, and Z. Guan, "Deeply understanding graph-based Sybil detection techniques via empirical analysis on graph processing," *China Communications*, vol. 17, no. 10, pp. 82–96, 2020.

[6] L. Nie, D. Jiang, and Z. Lv, "Modeling network traffic for traffic matrix estimation and anomaly detection based on Bayesian network in cloud computing networks," *Annales des Telecommunications*, vol. 72, no. 5–6, pp. 297–305, 2017.

[7] M. Mardani and G. B. Giannakis, "Estimating traffic and anomaly maps via network tomography," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1533–1547, 2016.

[8] N. Etemadi Rad, Y. Ephraim, and B. L. Mark, "Delay network tomography using a partially observable bivariate Markov chain," *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 126–138, 2017.

[9] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic matrix estimation: existing techniques and new directions," in *Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications - SIGCOMM '02*, p. 161, Pittsburgh, Pennsylvania, USA, 2002.

[10] D. Daniel-Simion and G. Dan-Horia, "Traffic shaping and traffic policing impacts on aggregate traffic behaviour in high speed networks," in *2011 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pp. 465–467, Timisoara, May 2011.

[11] D. Jiang, Z. Zhao, Z. Xu, C. Yao, and H. Xu, "How to reconstruct end-to-end traffic based on time-frequency analysis and artificial neural network," *AEU - International Journal of Electronics and Communications*, vol. 68, no. 10, pp. 915–925, 2014.

[12] F. Qian, G. Hu, and J. Xie, "A recurrent neural network approach to traffic matrix tracking using partial measurements," in *2008 3rd IEEE Conference on Industrial Electronics and Applications*, pp. 1640–1643, Singapore, June 2008.

[13] M. Emami, R. Akbari, R. Javidan, and A. Zamani, "A new approach for traffic matrix estimation in high load computer networks based on graph embedding and convolutional neural network," *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 6, 2019.

[14] K. Rajawat, E. Dall'Anese, and G. B. Giannakis, "Dynamic network kriging," in *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 121–124, Ann Arbor, MI, USA, Aug. 2012.

[15] K. Rajawat, E. Dall'Anese, and G. B. Giannakis, "Dynamic network delay cartography," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2910–2920, 2014.

[16] L. Nie, "A novel network tomography approach for traffic matrix estimation problem in large-scale IP backbone networks," in *2015 International Conference on Computer Science and Mechanical Automation (CSMA)*, pp. 97–101, Hangzhou, China, Oct. 2015.

[17] M. Coates, Y. Pointurier, and M. Rabbat, "Compressed network monitoring for ip and all-optical networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement - IMC '07*, p. 241, San Diego, California, USA, 2007.

[18] S. Qazi, S. M. Atif, and M. B. Kadri, "A novel compressed sensing technique for traffic matrix estimation of software defined cloud networks," *KSII Transactions on Internet and Information Systems*, vol. 12, no. 10, 2018.

[19] P. Tune and M. Roughan, "Spatiotemporal traffic matrix synthesis," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, pp. 579–592, London United Kingdom, Aug. 2015.

[20] A. Soule, A. Nucci, R. Cruz, E. Leonardi, and N. Taft, "How to identify and estimate the largest traffic matrix elements in a dynamic environment," *ACM SIGMETRICS Performance Evaluation Review*, vol. 32, no. 1, pp. 73–84, 2004.

[21] C. Tebaldi and M. West, "Bayesian inference on network traffic using link count data: rejoinder," *Journal of the American Statistical Association*, vol. 93, no. 442, p. 576, 1998.

[22] Y. Vardi, "Network tomography: estimating source-destination traffic intensities from link data," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 365–377, 1996.

[23] D. B. Chua, E. D. Kolaczyk, and M. Crovella, "Network Kriging," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 12, pp. 2263–2272, 2006.

[24] X. Fan and X. Li, "Network tomography via sparse Bayesian learning," *IEEE Communications Letters*, vol. 21, no. 4, pp. 781–784, 2017.

[25] L. Nie and D. Jiang, "A compressive sensing-based network tomography approach to estimating origin-destination flow traffic in large-scale backbone networks," *International Journal of Communication Systems*, vol. 28, no. 5, pp. 889–900, 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 3860083, 13 pages
https://doi.org/10.1155/2021/3860083

Research Article

Intelligent Link Prediction Management Based on Community Discovery and User Behavior Preference in Online Social Networks

Jun Ge,[1,2,3] **Lei-lei Shi**,[1,3] **Lu Liu**,[4] **Hongwei Shi**,[2] **and John Panneerselvam**[4]

[1]*School of Computer Science and Telecommunication Engineering, Jiangsu University, China*
[2]*School of Information Engineering, Suqian University, China*
[3]*Jiangsu Key Laboratory of Security Tech. for Industrial Cyberspace, Jiangsu University, China*
[4]*School of Informatics, University of Leicester, UK*

Correspondence should be addressed to Lu Liu; l.liu@leicester.ac.uk

Received 7 April 2021; Accepted 19 May 2021; Published 1 June 2021

Academic Editor: Varun Menon

Link prediction in online social networks intends to predict users who are yet to establish their network of friends, with the motivation of offering friend recommendation based on the current network structure and the attributes of nodes. However, many existing link prediction methods do not consider important information such as community characteristics, text information, and growth mechanism. In this paper, we propose an intelligent data management mechanism based on relationship strength according to the characteristics of social networks for achieving a reliable prediction in online social networks. Secondly, by considering the network structure attributes and interest preference of users as important factors affecting the link prediction process in online social networks, we propose further improvements in the prediction process by designing a friend recommendation model with a novel incorporation of the relationship information and interest preference characteristics of users into the community detection algorithm. Finally, extensive experiments conducted on a Twitter dataset demonstrate the effectiveness of our proposed models in both dynamic community detection and link prediction.

1. Introduction

With the rapid development of social networks and the wide spread application of intelligent terminals, we are facing an unprecedented volume of data generation, which in essence referred to as the big data era [1–3]. The big data era has led to various changes in the society and in our everyday lifestyle, where social networking is playing a pivotal role with great significance [4], which is a field of managing large-size datasets in a distributed computing environment. These datasets require robust network algorithms to transport huge block files efficiently. The traditional processing dataset approach involves a basic data placement technique that delivers resultant data blocks and exchanges block replicas in the cluster. And the widespread use of the Internet around the world enables people to connect with each other. People use the Internet for various purposes such as watching movies, reading news, searching information via search engines, shopping in e-commerce websites, and establish connections with friends via online social networks [5–8].

Online social networks are now flooded with various forms of data in massive volume, as people make new connections, post their updates, share and comment on other updates, etc., and such a structure of online social networks emphasizes the need to study their social relationships [9]. Besides, users can obtain information from related objects or spread information. Predicting the possible links between objects and/or information, based on the known information [10], would enable us to better understand the evolution of social networks [11, 12], also help business planners to make decisions, carry out precise services based on user connections, and achieve greater business value [4, 5, 9, 13–15].

In recent years, the essence of online social networks is largely identified to reflect the real-world networks. In the

Internet, people create their own content, tag, like, comment or join the community, and connect with other users. Better recommendation of potential friends for users can increase the connectedness of users with a better user experience. So far, recommendation systems are widely used in online social networks [16, 17], for various purposes such as new friend suggestions, and to offer efficient information retrieval services, etc., which in turn increase the overall traffic flow in the Internet.

In this context, the recommendation system used for suggesting friends introduced into the social network platform forms the basic function of social network services. The recommendation technology based on link prediction has become a research hotspot in recent years because of its high accuracy and low algorithm complexity. Link prediction is one of the basic problems in social network analysis, resolving which can provide us with an understanding of the mechanism of network evolution in theory, and further help us to optimize social networking services in accordance with the evolution of network structure. The concept of link prediction involves utilizing network structure information and node attributes to predict the possibility that users who have not yet generated link relationship in the network becoming friends in the future, to recommend the results with high possibility as "target users," so it is naturally suitable for relationship recommendation in social networks.

Although we can learn from the research on recommendation methods in traditional social network, due to the complexity and diversity of social networks, there are still many problems yet to be resolved in the field of relationship recommendation in online social networks, and the accuracy of relationship recommendation needs to be further improved [9, 18, 19]. To this end, we intend to develop a more effective relationship recommendation algorithm that is suitable for online social networks, characterizing higher prediction accuracy, low algorithm complexity, and better system integration. Herein, we propose an improved intelligent link prediction management technique based on exploiting the relationship strength information in online social networks. Moreover, considering the network structure attributes and interest preference of users as important factors affecting the links, further improvement is achieved by encompassing the community detection algorithm with the relationship information and interest preference characteristics of users. Finally, this paper designs a friend recommendation model, by integrating the intelligent link prediction algorithm and the label propagation community detection algorithm.

The main contributions of this paper are as follows:

(1) This paper proposes a community detection algorithm-based user behavior preference model (UBP), which can improve the data quality from the source of community detection. Specifically, in the calculation of community influence on nodes, the influence from nonadjacent nodes is also included. Through multiple iterations of experiments, the proportion of influence weight between adjacent nodes and nonadjacent nodes is identified to conform to real environment. Extensive experiments show that the proposed community detection results are better than the existing methods, and our community detection structure is more reasonable and accurate. Based on the UBP algorithm, the DPRank algorithm is introduced [4], where the global influence is replaced by the topology of social network and the local influence of nodes. Our approach not only ensures the accuracy of the algorithm but also improves its efficiency

(2) This paper proposes the novel link prediction algorithm based on label propagation. Firstly, we collect the attribute features and text information of users to explore their potential preferences and extract tags and then construct the user feature vector model to calculate the similarity between users. Then, based on an improved multisource label propagation community detection algorithm (multisource label propagation algorithm (MSLPA)), similar communities are mined. Finally, based on community, we use link prediction to estimate the node pairs with closest relationship strength and select the Top-k potential friend list as recommendation to users. This method not only improves the accuracy but also reduces the computational complexity of the link prediction algorithm. Performance evaluation carried out based on the real dataset shows that our algorithm achieves better performance than the state-of-the-art local index methods

(3) We conducted experiments to evaluate the performance of our proposed models. The experimental results on a Twitter dataset demonstrate the effectiveness of our proposed UBP and MSLPA models, in terms of both dynamic community detection and link prediction

The rest of this paper is organized as follows. In Section 2, we review previous studies of link prediction. In Section 3, we introduce our proposed UBP method. We present the MSLPA model in Section 4. We discuss our experimental results in Section 5, and in Section 5.1, we draw our conclusions and future work.

2. Related Work

Online social networks focus on the interaction between individuals and network topology. Internet, scientist cooperation network, power network, aviation network, biological network, and so on, all reflect the characteristics of social networks [4, 12]. It is worthy of note that most of the interconnected things can be abstracted as social networks. Typical networks, such as the cooperative network between scientists in the academic field, and the network structure of protein molecules in the biological field [18–20] resemble the topology of social networks. The ultimate purpose of studying the topological structure and properties of different types of networks is to understand the evolution principle of social networks, predict the future evolution direction and trend of social networks, estimate links [21] to better cope with the sudden changes in social networks, and apply this

knowledge in actual networks [22–24]. For example, in the field of counterterrorism, law enforcement officers can analyze social network links to identify the direct and indirect connections of suspects. "Guess what you like" in e-commerce website, recommendation of the target of interest in Twitter [25], etc., can be seen as the application of link prediction in real life. Because of its significant practical value, link prediction and recommendation systems have become the hotspot research topics in the context of online social networks [26, 27].

As one of the important research directions of data mining, link prediction in online social networks has received a wider attention, and many link prediction algorithms have been developed in the recent years. Traditional research methods [22, 28, 29] have two main ideas: Firstly, from the perspectives of machine learning, link prediction modeled as a typical learning problem and witnessed the application of techniques such as supervised logistic regression, support vector machine, random forest, and unsupervised learning algorithm based on Bayesian network [24]. The other idea is to mine the properties of nodes and network structure from the perspectives of social networks and predict the connection based on the similarity of nodes. These methods attempted to mine node and network related information as much as possible. Moreover, the maximum likelihood method is also heavily researched for link prediction and witnessed to have achieved reliable prediction results.

Srinivas et al. [25] comprehensively analyzed the importance of link prediction for social network analysis along with its application in bioinformatics, information retrieval, e-commerce, and other fields and summarized various link prediction technologies based on classification and kernel function and discussed the latest progress and future research direction of probability modeling in this context. Yang et al. [28] compared the advantages and disadvantages of various link prediction algorithms and conducted quantitative research in real networks. They pointed out that the similarity method based on network topology has become a research hotspot due to its simple algorithm and low computational complexity. According to the influence of different nodes, an improved similarity link prediction algorithm is proposed.

Li et al. [23] defined a preference function as a new attribute of supervised learning considering the preference of nodes and achieved reliable results. Gupta and others [24] abstracted the link prediction problem into a binary problem and established a Bayesian model to predict the possible connections of the network. The link prediction method based on probability model has been applied to various fields of social network research, in an attempt to establish a perfect social network recommendation system [30].

Bastami et al. [27] believed that most of the link prediction algorithms only consider either global or local information, but only few of them integrate both global and local structure information. A new fusion algorithm has been proposed, which considers community properties into account, and used the clustering algorithm to evaluate the link density at the community level to adjust the eigenvalues of nonlocal features and then combined the link information and nodes of neighbor nodes. Finally, the similarity model has been integrated to predict the link. This algorithm innovatively integrated the local features of nodes and the structural features of communities.

Community detection and link prediction are two different directions of social network analysis [31], while the former is used to mine network topology, the latter works based on the network structure to predict the future evolution trend in the social network. At present, a few researchers have tried to use community detection to improve the accuracy of link prediction. Yao et al. [32] studied the significance of clustering coefficient in link prediction and proposed a new periodic evolution model. Experiments on the Enron network dataset showed that the prediction ability of this model is better than that of the classical link prediction algorithms.

Link prediction and community detection are of great significance in social network analysis, as they describe the formation, evolution, and nature of the social network from different aspects. In this paper, we propose a new link prediction method for social networks by incorporating community detection, ultimately to offer valuable friends recommendation for users.

3. Concept and Definition

3.1. Social Network. In the real world, there is a wide range of connections and interactions between various things. The components of the system can be described as nodes. Many of these systems can be modeled by social networks. Studying the formation mechanism and evolution mechanism of social networks can help us to understand the nature of the system better. For example, the discovery of six-degree separation theory in a typical social network shows the world is actually very small. Link prediction involves solving one of the most fundamental problems in network science, which is to restore and predict the missing information. When a system becomes more complex, many nodes that have not been linked at present may establish links in the future. The problem of predicting such nodes with a higher likelihood of establishing links in the future is called link prediction [33]. The interaction or connection between nodes is described as the edge between nodes. A typical social network usually characterizes frequent connections between nodes, where new links become more active.

Any network can be abstracted as a graph, which is composed of finite sets. Such a graph structure encompasses node set representing the individual in the network, edge set representing the connection in the network. Generally, a network can be represented as a node or entity set of the same type. A social network usually encompasses a specific user, a link set, and a link between nodes. If a node has a complete set of possible links, the nonexistent link instances can be represented as a prediction problem of generating links. Thus, the link prediction problem can be defined as follows: given an instance of a social network, we can predict the possibility of generating links and judge the possibility of connectedness according to the score value. Generally, such a kind of prediction problem is studied with a training set and a test set.

Figure 1(a) represents a complete network, which consists of 12 nodes and 16 edges. Four edges are extracted as test

edges as shown in Figure 1(b), and the remaining 12 edges are training datasets. Through a link prediction algorithm, four test edges are given a score value according to the possibility and compared with all other edge scores that do not exist. A higher score value reflects a more accurate prediction result.

3.2. Community Structure. Community is a subgraph structure in the network topology; as shown in Figure 2, the density of node links within the community structure is higher than that of between communities. This implies that the internal relationship within communities is closer and in line with the cognition of real-world social communities.

4. UBP Model

In this section, we describe our proposed community detection algorithm based on user behavior preference (UBP), which can improve the data quality from the source of community detection. In the calculation of community influence on nodes, the influence from nonadjacent nodes is also included. Through multiple iterations of experiments, the proportion of influence weight between adjacent nodes and nonadjacent nodes is identified to conform to the real environment. Based on the UBP algorithm, the DPRank algorithm is introduced [4], where the global influence is replaced by the topology of the social network and the local influence of nodes, which ensures the accuracy of the algorithm and improves its efficiency.

4.1. Community Detection. In this section, we explain the community detection method.

As demonstrated in Figure 3, we divide the social network into communities and determine the exact community structure. We propose the UBP algorithm to achieve the desired objectives. In the UBP algorithm, we consider the topological relationship between the adjacent users in the community, the topological relationship between indirect users, and the impact probability between the users. The UBP algorithm provides a good community structure. Then, in the divided community structure, we use the IPIP model [8] based on the LT model to determine the most effective nodes in each community in order to maximize the impact of the social networks. Finally, a situation where a possible loss of seed nodes can occur in the real society, we use the Full Preselected Search, namely, FPSS algorithm [8]. When the seed node is lost, the FPSS algorithm immediately finds a replacement node to compensate for the loss of influence diffusion caused by the loss of the original seed node.

Besides, we pursue the following steps [8], as illustrated in Figure 4.

(1) The user interest is modeled, and the Pearson coefficient is used to obtain the interest similarity matrix between users

(2) The influence probability of users is modeled based on their concerns and interactions

(3) Social network is modeled based on user interest similarity and user influence probability

(4) All the users are calculated and sorted in the social graph accordingly, where the central community and the central nodes are identified, and finally the community is detected

(5) Link prediction based on independent cascade model is achieved

(6) Link prediction in the entire network is achieved

4.2. Modeling User Interest Similarity

4.2.1. Modeling User Interest. Users share their feelings and thoughts in Twitter by posting a tweet and participating in social activities. The LDA model [4, 8] is a three-level Bayesian model of document subject word, which uses probability deduction to find the semantic structure of a given dataset to obtain the topic of the text. Hence, the LDA algorithm is used to analyze the user's document for obtaining the user-interest matrix. On this basis, the interest similarity between users in social networks is resolved.

4.2.2. Modeling User Similarity. A substantial amount of hidden information is present in the massive data. We can use this gigantic data to extract the desired information by analyzing the data and then receiving the user score on posts and finally generating the user-post matrix. Pearson's sparse is an efficient technique for obtaining user similarity. Pearson's coefficients given in equation (1) with a modified cosine similarity and user-post scores are used to compute the similarity of users in the network.

$$w_{\text{pearson}}(i,j) = \frac{\sum_{u \in I(i,j)} \left(r_{i,u} - \overline{r_i}\right)\left(r_{j,u} - \overline{r_j}\right)}{\sqrt{\sum_{u \in I(i,j)} \left(r_{i,u} - \overline{r_i}\right)^2} \sqrt{\sum_{u \in I(i,j)} \left(r_{j,u} - \overline{r_j}\right)^2}}, \tag{1}$$

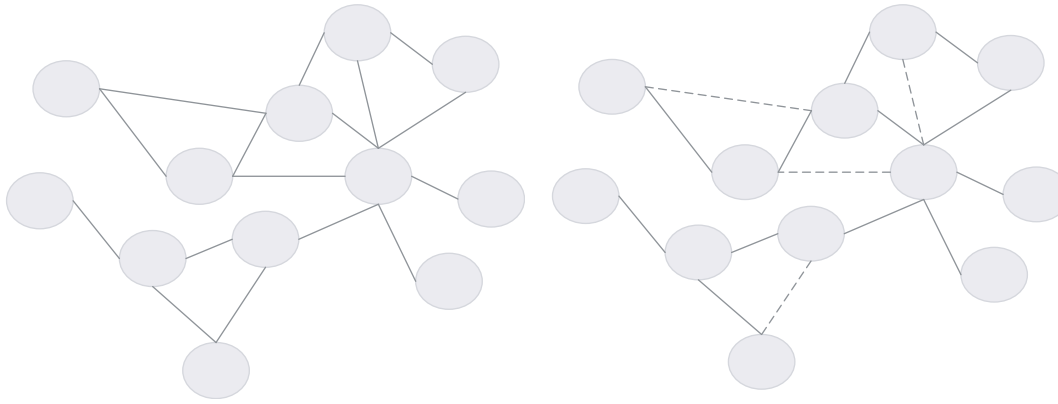
where $I(i,j)$ shows the set of common posts of user u and user v. While $r_{i,u}$ represents the score of user i on post u, and $\overline{r_i}$ denotes the average interest scores of user i and user j.

4.3. Modeling the User Influence Probability

4.3.1. Initial Influence Probability Modeling for Users. In general, user's influence is the degree of trust between each other while the community's influence on the average user is the sum of all the influence in the community.

In this paper, the influence probability of users originates from the number of interactions between users, which mostly reflects the influence of the relationship between the users. Thus, we evaluate the initial influence probability based on the user's interactions. The initial influence probability of user u on user v is calculated using

$$F(u,v) = f(u,v), \tag{2}$$



(a) Complete network (b) Prediction network

Figure 1: An example of link prediction.

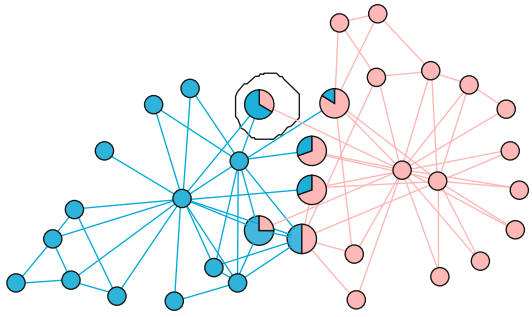


Figure 2: Community structure.

where $f(u, v)$ shows the number of interactions between users u and v.

4.3.2. Probability Prediction of Influence between Unconnected Users. In social networks, users have almost no explicit source of influence. A study on social networks [5] shows that the association between two unfamiliar users' average 4.7 hops, i.e., edges. By integrating research and reality, users will have more trust on their friends of friends. We use equation (3) to propose the probabilistic modeling for source nodes and two-hop target nodes.

$$P(A, C) = \begin{cases} F(A, C), A \to C \\ \dfrac{1}{|S|} \sum_{j \in S} P_{AB_j} P_{B_j C} + F(A, C), A \to B_j \to C \end{cases}, \tag{3}$$

where $A \to C$ indicates that user A is the follower of user C. S is a collection of common friends of users A and C, and P represents the probability of a user's influence on another user.

Figure 5 illustrates a graphical representation of the influence of a probability between the users. Moreover, it illustrates the relationship between two nodes in the network, where the weight value denotes the probability that a user may receive from another user.

Figure 6 shows the link relationship between many nodes, where black solid lines represent the concerns among the users while red dotted lines indicate the probability of predicting the influence between them. At this point, we believe that the probability of influence between users with links is obtained. We represent this social graph in the form of a matrix. For example, the processing of calculating $P(U_1, U_4)$ is given as follows:

$$P(U_1, U_4) = \frac{1}{2}(P(U_1, U_2) * P(U_2, U_4) + P(U_1, U_3) * P(U_3, U_4)) + F(U_1, U_4) \approx 0.678. \tag{4}$$

4.4. Modeling Social Networks. We model the social network based on user influence probability and user interest similarity. The weight of the side in a social network can be calculated using

$$F(i, j) = \eta P(i, j) + (1 - \eta) w_{\text{pearson}}(i, j), \tag{5}$$

where η is a tunable parameter that regulates the importance of user influence probability and user similarity in the social network model. After the experimental observations, we set $\eta = 0.4$. The result of equation (5) is the weighting of a social network. This is because the user influence is not equal and produces a realistic response.

5. MSLPA Model

Preference connection has proved to be an idea that can improve the accuracy of link prediction. On this basis, this paper proposes an integration of the label propagation and the link prediction algorithm. Firstly, we collect the attribute features and text information of users to explore their potential preferences and extract tags and then construct the user feature vector model to calculate the similarity between users. Then, based on an improved multilabel propagation community detection algorithm (multisource label propagation algorithm (MSLPA)), similar communities are mined. Finally, based on community, we use link prediction to identify the node pairs with the closest relationship strength and select

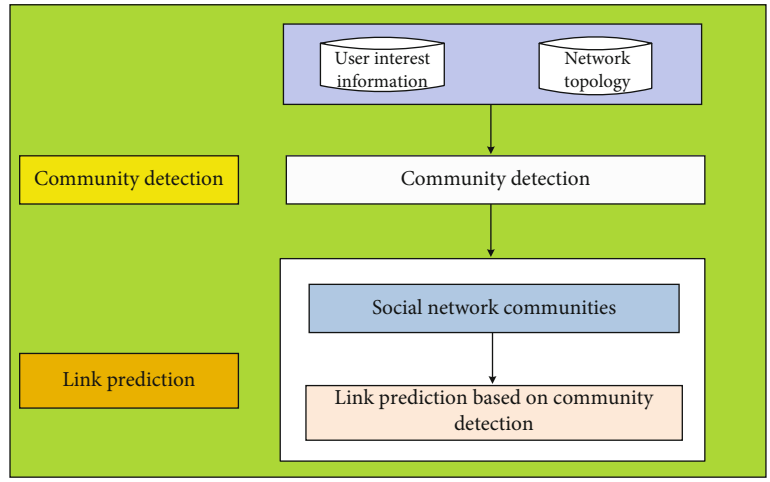


Figure 3: Link prediction method based on community detection.

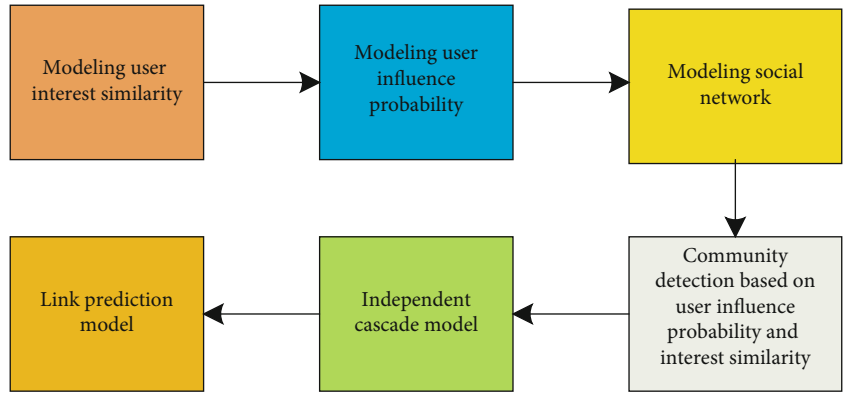


Figure 4: UBP algorithm steps.

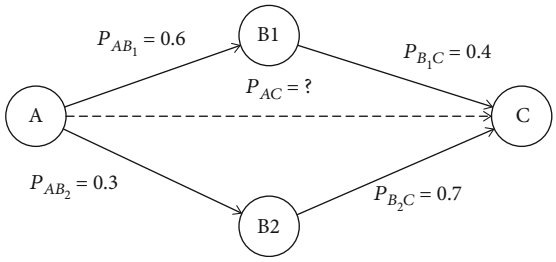


Figure 5: Schematic representation of user influence probability prediction.

the Top-k potential friend list as a recommendation to users. This method not only improves the accuracy but also reduces the computational complexity of the link prediction algorithm.

In this section, we first introduce the label propagation model and then propose a link prediction recommendation algorithm combined with label propagation. Unlike the single similarity index, in combination with the relationship strength, our model takes into account not only the local index but also the global structure information and user attribute information. In comparison with the traditional link prediction algorithm, our proposed model not only improves the accuracy of recommendation but also makes the network more compact after the label propagation. Furthermore, the required amount of calculation is considerably reduced, as our model avoids computation for all the nodes; thus, it is fast and efficient.

5.1. Link Prediction. Community detection can be used to mine the social network user information and network structure attributes and further can be applied for link prediction. User information and network structure usually complement each other well. Social network theory shows that people with similar characteristics tend to establish a relationship. Traditional link prediction methods are required to calculate information about all the nodes in the whole network. Due to its high computational complexity, this strategy incurs a large amount of calculation, which significantly affects the efficiency of the prediction algorithm in large-scale social networks. In order to solve this problem, researchers put forward the idea of dividing the large-scale social network into communities and then using link prediction to calculate the similarity within the communities. For the recommendation system, it is equivalent to the recall stage first, which not only reduces the computation scale but also makes the recommendation source more targeted. This paper introduces

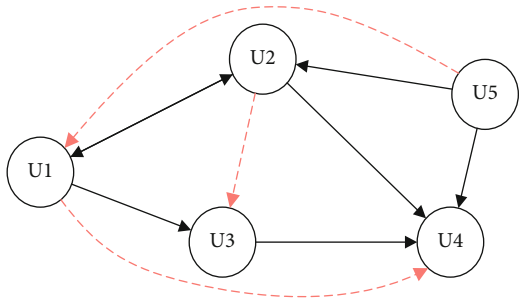


Figure 6: Schematic diagram of links between the nodes.

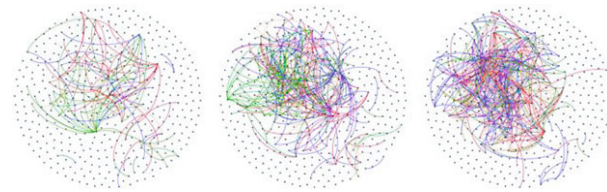


Figure 7: Community evolution.

an improved multilabel propagation algorithm. Tag extraction uses user preference features extracted from user self-set and topic model, divides the community, and then calculates the similarity through link prediction. According to the similarity ranking, it further helps users to find friends who have similar interests, or they may know. In this way, our recommendation algorithm makes use of the user's personal preference attributes and social network structure information to make the recommendation more personalized and improve the recommendation accuracy.

5.2. Community Detection. With the gradual deepening of social network research, people begin to realize the existence of locally connected node sets in the network, which have a very important impact on the topological structure of the whole graph.

Community detection involves mining the communities in the network through various algorithms, so as to analyze the communities and understand the evolution trend of communities. Figure 7 shows the evolution process of different types of communities on a social platform. The community detection algorithm based on label propagation algorithm (LPA) is efficient and simple, with only linear complexity. It is suitable for large-scale networks and widely used in industry.

Tags reflect the interests and characteristics of users, and the process of tag propagation reflects the simulation of human information exchange. The process of extracting tags from node behaviors for further propagation retains the node's personality. Based on the nodes after tag propagation, some communities based on similar interests are formed. This idea is similar to the process of community formation in social networks, where users tend to join most of the neighbor's community. With the gradual expansion of social networks, various types of communities are formed. In this paper, some local similarity indexes such as CN, Jaccard, and AA are compared based on tags. The results show that these algorithms can achieve good prediction results in the case of relatively dense data. Community structure exists objectively, but users in a certain community only interact with those users who have a direct connection with them. However, in a community, users who are not directly connected are also regarded as being "close." Friend recommendation system usually gives priority to users belonging to the same community, as birds of a feather flock together, which reflects the community detection algorithm. In fact, it divides the community in a certain way. On this basis, it can make further link prediction for each community. From the computational perspective, community detection can be regarded as equivalent to the decomposition of the network, which reduces the computational complexity.

At present, information on social networks is growing explosively. Tag propagation algorithm has become a research hotspot because of its simplicity, high efficiency, and low computational complexity. In order to facilitate label management, we establish corresponding tag systems, which can help users to retrieve users and related resources they are interested in.

The traditional label propagation algorithm assumes that each node only carries one tag, belongs to a community, and does not distinguish the importance of different tags, which is not consistent with the real-world social networks. This paper improves the tag propagation algorithm, distinguishes the tag weight, and can carry out multilabel propagation. Finally, those nodes with the maximum similarity believe to be in the same community but actually may belong to multiple communities at the same time.

5.3. Label Selection and Extraction. Users often set up some personal tags or post some blog posts and comments in social networks. The analysis and mining of this text information make the recommendation more personalized. In the Twitter network, users usually add their own tags to reflect their personality.

The main idea of label propagation algorithm is described as follows: suppose a node "a" and its neighbor nodes are $\{A_1, A_2, A_n\}$. Each node carries its own label. During the process of propagation, the label of node "a" is determined by the label of its neighbor node; that is, the label of node "a" with most neighbors is taken as the label of node "a." With the continuous spread of tags, it tends to be stable in the end.

The traditional LPA algorithm does not distinguish the importance of tags. In real networks, there are often first-class, second-class, and third-class tags. For example, in the user profile system, the tags selected by users themselves are more important than the tags that are counted out. In order to make the community detection more reasonable, we distinguish according to the importance of labels. In this paper, we mainly determine the following two types of labels with different weights and adjust the weights of various labels through experiments according to the actual situation.

(1) Labels and system tags set by users themselves

(2) There are obvious tags in the user text, such as ∗∗ school and ∗∗ location

In addition, we use the LDA topic model to extract the corresponding tags according to the blog information published by users. In the tag extraction process, we input some candidate seed words. Since the entity set is not directly represented in the user's Tweets, we need to use specific tools to find the candidate entity set and then compare the similarity with the seed vocabulary to obtain the required classification tag. The input document of the model is collected based on historical behavior data of users. Finally, it is merged with the first two types of tags as a user's tag set. Users in social networks generally have multiple tags. It is obviously not suitable for social networks to select only one tag in the propagation process, as in the case of the traditional tag propagation algorithm. In this paper, we propose an improved multisource label propagation algorithm (MSLPA).

5.4. Improved Label Propagation Algorithm. Considering the interaction characteristics of real social networks, we improve the classic label propagation algorithm and apply it to the whole recommendation system. The improved communication process can be divided into the following three steps:

Step 1. Initialize labels instead of community numbers for all nodes, and nodes carry multiple labels, and assign weights to labels at the same time.

Step 2. Refresh the labels of all nodes iteratively. The label of all neighbor nodes is investigated, and the weight is calculated; then, the labels are assigned with the largest number to the current node. When the number of labels with the largest number is not unique, select one randomly.

Step 3. After n iterations, convergence is reached, and the algorithm is completed. In the final community, the nodes with the greatest similarity degree belong to the same community.

Considering the computational complexity, the most widely used text-matching model in the field of text analysis is vector space model (VSM) [13]. In the concept of VSM, a document is represented in vector form, and its relevance is measured by the similarity between vectors. Each dimension of the vector corresponds to a term, and the weight of each component element of the vector represents the importance of the term in the document. This paper studies the label words, which can be abstracted as document vectors. The way to calculate semantic similarity using cosine similarity can be expressed as follows:

$$\mathrm{RSV}(A,B) = \frac{\sum_{i=1}^{n}(a_i \times b_i)}{\sqrt{\sum_{i=1}^{n}(a_i)^2} \times \sqrt{\sum_{i=1}^{n}(b_i)^2}}. \tag{6}$$

When the value of $\mathrm{RSV}(A,B)$is 1, it means that the labels between the two nodes are the same; when the value is 0, there is no overlapping label between the two nodes. All neighbor nodes whose values exceed the predetermined propagation threshold are selected, and the most selected labels are passed to the corresponding nodes. Finally, we use the improved link prediction algorithm to predict the possible edges in the community.

The propagation process is shown in Figure 8. The label propagation process simulates the information exchange process and behavior of people in social networks. During initialization, each node has a fixed label, which is uploaded by the user and extracted from the previous LDA model. Then, some nodes are randomly selected to interact with other nodes. The $\{A, B, C, D\}$ in Figure 8(a) represents the label currently owned by the node. We randomly select the node {4, 2, 1} as the receiving node. First, node 4 receives the label from neighbor 1 and the label from neighbor 2, and the label of node 4 becomes $\{D, A, B\}$.Then, the label of node 2 is updated. Since its neighbor node 3 has multiple labels, a label is randomly selected and passed on to node 2. Here, suppose the propagation label is selected, and the label of node 2 becomes $\{B, A, C, D\}$. Finally, the label of node 1 is updated. The neighbor nodes {2, 3, 4} are randomly selected to propagate. If there are repeated labels, the corresponding label weight is increased by one in turn, for example, after the propagation in Figure 8(b) is finished. If there are two "a" labels in node 1, the weight is two, and the weight of $\{B, C\}$ labels is one. The whole propagation process is asynchronous. Some nodes will receive the label information of neighbor nodes first and can end the propagation process according to the label propagation. Finally, according to the different labels of stable nodes, the similarity is calculated by formula and divided into multiple communities. At the same time, a node can belong to multiple communities.

The time complexity of the algorithm is very low. In the process of label propagation, nodes are randomly reordered to ensure the convergence of the algorithm. Because the formation of community only depends on the local information of the network, the algorithm is very suitable for community detection and partition in large-scale social networks.

Firstly, the initial seed node is set, and the label is propagated to the surrounding nodes according to the label weight. After each round, the similarity between nodes is calculated according to formula (1), and then, the label is updated.

Finally, according to the specified number of iterations, the program ends after the community becomes stable.

Label classification and link prediction can promote each other and have homogeneity in social relations; that is to say, people with similar attribute characteristics are more likely to establish friendship relationship; therefore, the closer the relationship is, the more likely they are to have the same label.

The nodes in the circle in Figure 9 are communities formed based on a certain relationship. It can be seen that nodes I and E, M, and K are equally likely to generate links through calculation. However, since I and E belong to a certain community relationship, there are similarities between them. Therefore, link prediction based on label division is more likely to recommend friends within a community, hence can be more targeted. On the other hand, users' interests are

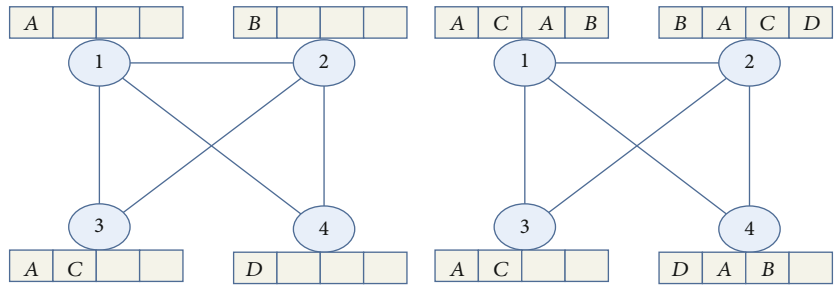


Figure 8: Schematic diagram of multisource label propagation process.

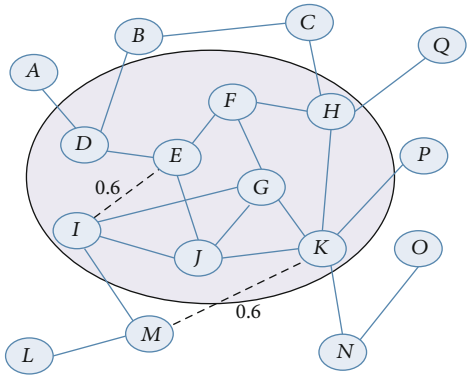


Figure 9: Schematic diagrams of link prediction differences among different nodes.

diverse, and users tend to add topics of interest and find similar people. Through the way of label propagation, more potential friends can appear in the recommendation list, which also reduces the cold start problem of new users to a certain extent.

Here, the first stage uses the multisource label propagation algorithm, as in algorithm one. The algorithm inputs the preprocessed network data, including adjacency matrix and label set. In the second stage, the MSLPA algorithm is used to calculate the similarity of the corresponding users, and the Top-k potential users are selected to generate the recommendation list.

5.5. Recommendation Model Based on Combination Algorithm. In this section, this paper proposes a friend recommendation system model by integrating label propagation and link prediction. The whole recommendation process is divided into four modules: data cleaning, community detection, link calculation, and user recommendation. The whole system model framework is shown in Figure 10.

Firstly, user's home page information and blog text information are collected, and a single microblogging is regarded as a short text. Then, the label is generated by extracting the subject words of the blog post content using the LDA model. Then, the ID information of the followers and their followers is obtained, and the labels are saved in the corresponding table after being extracted for further data cleaning.

Secondly, we read the obtained label and link relationship data, use the community algorithm to divide the community, and stop the iteration when the community is relatively stable to the set conditions.

Finally, we traverse all the communities formed in the previous step, use the improved link prediction algorithm to calculate the similarity within the community, and select the Top-k users as the recommendation list for each user according to the score ranking. For the friend recommendation for a given user, the community to which the user belongs to is first obtained, and the potential friends with the highest similarity are ranked according to the final total score.

6. Experiments

In this section, we present the results obtained in our experiments conducted on real-world short-text data collections in order to demonstrate the effectiveness of our proposed method. We consider four typical algorithms as our benchmark methods, namely, CPM [9], COPRA [18], LFM [19], and GCE [20]. We also introduce the collection of the dataset, experimental setup, analysis, the baseline approach, and the model evaluation.

6.1. Dataset.

6.2. Experimental Setup. The experiments are conducted in a machine with Intel I5 2.5 GHz CPU and 4G memory. The experiments use standardized mutual information, named NMI to measure the correlation between the community structure generated by the community detection algorithm and the standard community structure to evaluate the accuracy of the algorithm using equation (7). We use the overlapping modularity Qov to evaluate the network structure of overlapping communities in order to measure the quality of community detection as expressed in equations (7)–(9).

$$\mathrm{NMI}(X|Y) = 1 - \frac{1}{2}\left(H(X|Y)_{\mathrm{norm})} + H(Y|X)_{\mathrm{norm})}\right), \quad (7)$$

where X and Y represent the experimental community structure and the standard community structure, respectively. The higher the NMI value, the more similar the partition result is to the standard network structure and the higher the accuracy of the algorithm to partition the community.

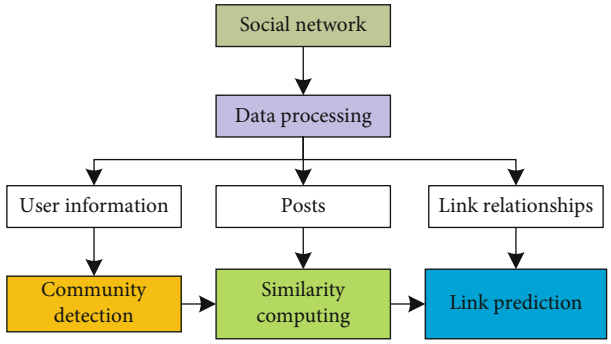


Figure 10: Framework of recommendation system.

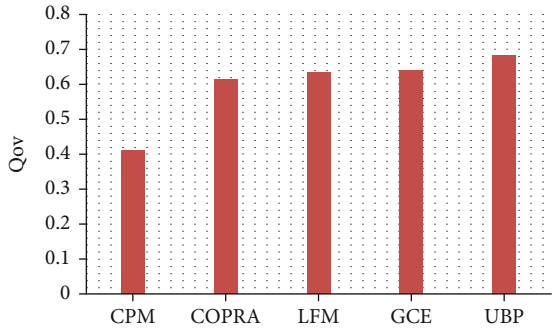


Figure 11: Comparison of Q values of algorithms for Twitter datasets.

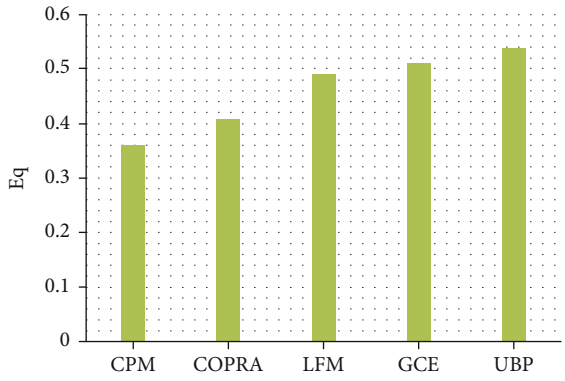


Figure 12: Comparison of EQ values.

Table 1: Test dataset detail table.

Network	Nodes	Edges	Coefficient	Degree
USAir	332	2126	0.749	12.81
NS	379	914	0.798	4.82
PB	1222	16714	0.360	27.36
Slavko	334	2218	0.488	13.28
Email	1133	5451	0.254	9.620
Router	5022	6258	0.033	2.49
Jazz	198	2742	0.618	27.70
Twitter	11241	732193	0.162	65.14

Table 2: UBP comparisons with other approaches.

		CPM	COPRA	LFM	GCE	UBP
Twitter	EQ	0.3545	0.4084	0.4902	0.5076	0.5387
	Q_{ov}	0.4122	0.6089	0.6298	0.6425	0.6856

$$\mathrm{Qov} = \frac{1}{m}\sum_{c\in C}\sum_{i,j\in V}\left[r_{ijc}A_{ij} - \omega_{ijc}\frac{k_i^{\mathrm{out}}k_j^{\mathrm{in}}}{m}\right], \tag{8}$$

where $A$is the adjacency matrix, K shows the degree of users, m indicates the number of edges, r_{ijc} denotes the probability that users i and j belong to community c, $r_{ijc} = \iota(P_{i,c}, P_{j,c})$, $P_{i,c}$ represents the probability that i belongs to community c, and ω_{ijc} denotes the probability that node i or node j belongs to community c.

$$\tau\left(P_{i,c}, P_{j,c}\right) = \frac{1}{\left(1+e^{-f\left(P_{i,c}\right)}\right)\left(1+e^{-f\left(P_{j,c}\right)}\right)}, \tag{9}$$

$$\omega_{ijc} = \frac{\sum_{j\in V}\tau\left(P_{i,c}, P_{j,c}\right)}{|V|} \times \frac{\sum_{i\in V}\tau\left(P_{i,c}, P_{j,c}\right)}{|V|}, \tag{10}$$

where f is defined as $f(x) = 60x - 30$ and Qov ranges from 0 to 1. The larger the value of Qov, the better the overlapping community structure will be.

COPRA, LFM, GCE, and CPM methods are selected comparative evaluation in the experiment. In order to avoid the influence of randomness of the algorithm in the experiments, we conduct 20 experiments and obtain the average results. Our model only needs one experiment because of the stability of the algorithm. The NMI and Qov of each algorithm are obtained. Figures 11 and 12 illustrates the changes in the NMI and Qov values in the social network with the mixing parameter.

6.3. Experimental Result. We test five algorithms on the Twitter dataset as shown in Table 1, which includes eight networks and their detailed information. Table 2 and Figure 11 demonstrate the comparison of community modules obtained after experiments on Twitter dataset, which terms the EQ and Qov with these algorithms, in which we can observed our proposed UBP algorithm performs better than the CPM, COPRA, LFM, and GCE algorithms due to the fact that our method considers both the user interest similarity and user influence probability with regard to Qov. And in the same way, Figure 12 shows the comparison of overlapping community modules attained from experiments on Twitter dataset. It can be observed that in the Twitter dataset, our algorithm also performs better than the CPM, COPRA, LFM, and GCE algorithms because our method considers both the user interest similarity and user influence probability.

Figures 11–14 demonstrate the output of our experimental analysis. Figures 13 and 14 show the effect of the variable parameter η and variable parameter λ, respectively.

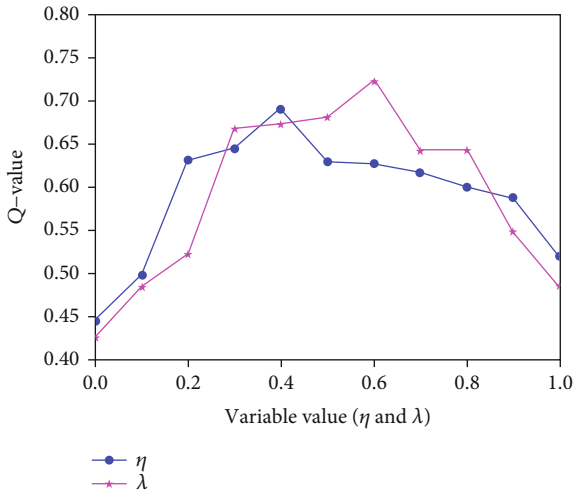


Figure 13: The effect of variable parameters.

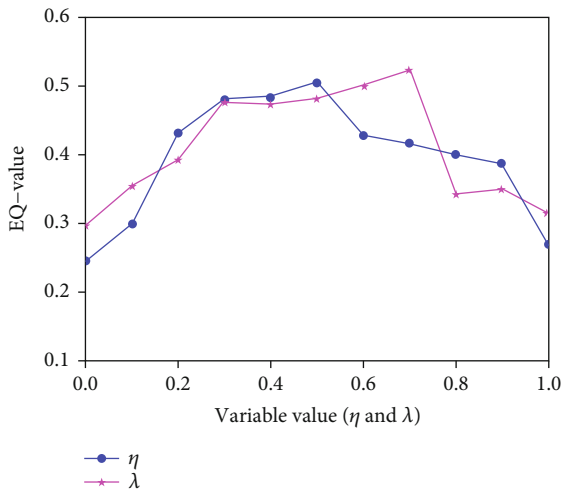


Figure 14: The effect of variable parameters.

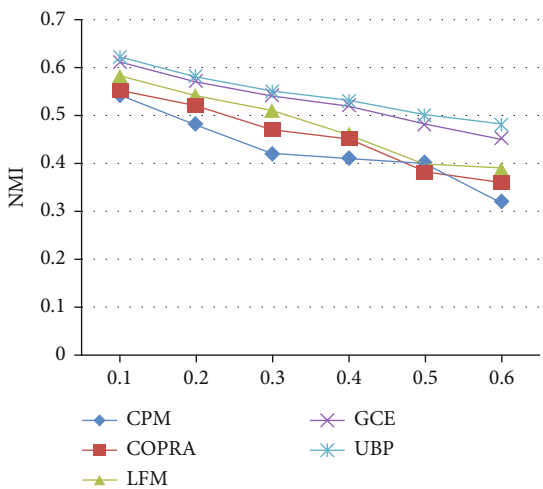


Figure 15: NMI comparisons of five algorithms.

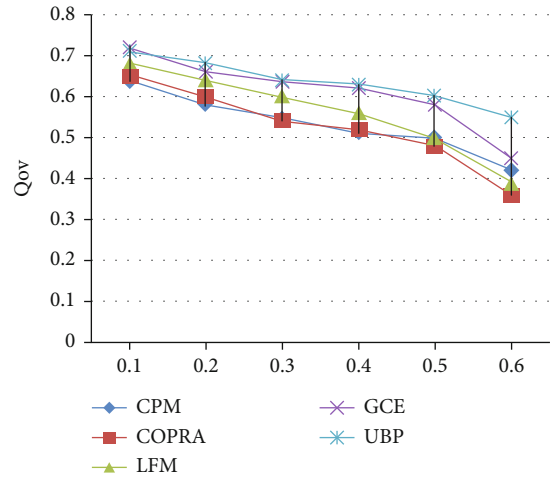


Figure 16: Qov value comparisons of five algorithms.

For variable weights and real weighted networks, we performed several experiments. The output values are not a variate during experimental analysis, and our algorithm shows an exceptional stability.

From Figure 15, when mixing parameter value is small, then the community structure of the network is more obvious, and the boundary between communities is clear. GCE characterizes a higher NMI value than UBP, but with an increase in the mixing parameter value, UBP algorithm exhibits a higher NMI value than the GCE algorithm. This shows the higher accuracy of the UBP algorithm in a large-scale network.

Figure 16 shows that the community modularity Qov divided by the GCE and UBP algorithms has more advantages than the other three algorithms regardless of the mixing parameter value. This is attributed to the improvement in the random strategy of the two algorithms, as it reduces the difference in the results and plays an important role for the users with higher potential influence to a certain extent. UBP has more potential influence, which can divide communities with high quality either in the network with obvious community structure or in fuzzy networks. The UBP algorithm improves the stability of community detection process and the quality of community generation to a certain extent, when compared with the existing community detection algorithms.

In order to verify the effectiveness of the proposed algorithm, link prediction algorithms such as CN, AA, RA, PA, JACCARD, HPI, LP, and Katz are selected for comparison. By comparing the results in Tables 3 and 4, it can be found that when compared with CN, JC, PA, AA, RA, HPI, LP, and KATZ algorithms, our proposed MSLPA algorithm delivers better prediction accuracy and AUC in most networks. In Table 3, the performance value of our MSLPA algorithm is improved by 10% to 20%, when compared with that of the traditional local index algorithms such as CN and RA. Furthermore, the average performance of PATH-based algorithms such as HPI and LP is also improved by 8.9%. In particular, our algorithm has more obvious advantages in the network with obvious social network properties, and its

Table 3: Precision comparison.

Networks	CN	AA	PA	RA	Jaccard	HPI	LP	Katz	MSLPA
USAir	0.628	0.722	0.657	0.690	0.494	0.677	0.742	0.641	0.755
NS	0.868	0.887	0.864	0.875	0.635	0.769	0.761	0.876	0.889
PB	0.816	0.833	0.858	0.827	0.667	0.850	0.739	0.733	0.904
Slavko	0.814	0.858	0.823	0.853	0.641	0.745	0.750	0.736	0.802
Email	0.845	0.867	0.859	0.811	0.643	0.813	0.816	0.819	0.844
Router	0.649	0.673	0.743	0.658	0.352	0.852	0.856	0.874	0.841
Jazz	0.775	0.802	0.796	0.772	0.361	0.798	0.811	0.802	0.829
Twitter	0.877	0.854	0.867	0.839	0.857	0.851	0.753	0.855	0.938

Table 4: AUC comparison.

Networks	CN	AA	PA	RA	Jaccard	HPI	LP	Katz	MSLPA
USAir	0.941	0.952	0.960	0.968	0.919	0.877	0.952	0.945	0.970
NS	0.968	0.980	0.964	0.975	0.976	0.979	0.981	0.983	0.988
PB	0.916	0.933	0.899	0.927	0.855	0.870	0.919	0.931	0.942
Slavko	0.954	0.948	0.939	0.953	0.946	0.945	0.950	0.936	0.959
Email	0.845	0.876	0.869	0.891	0.842	0.856	0.866	0.898	0.899
Router	0.649	0.677	0.943	0.658	0.651	0.652	0.946	0.957	0.660
Jazz	0.955	0.958	0.969	0.972	0.961	0.949	0.953	0.963	0.975
Twitter	0.971	0.966	0.977	0.963	0.957	0.951	0.952	0.965	0.982

prediction accuracy in PB network, Slavko network, and Twitter network is greatly improved. The clustering coefficient and average network degree of these networks are relatively large, the community structure is more obvious, and they are more likely to be connected by some relationship attributes. For example, the Jazz network, which consists of small groups of music, has a close relationship with a class of students on the Twitter social network. For Router, USAir, and other networks with weak social properties and sparse data, our MSLPA algorithm plays a very limited role in strengthening the relationship, when compared with other algorithms, and the prediction accuracy is not significantly improved.

In social networks such as Twitter, different mutual friends represent different relationship, which can be observed through the closeness of their neighbors. For this reason, our improved predictors have a good predictive effect on social networks. However, there is a lack of applicability for new users due to the lack of link information.

7. Conclusions and Future Work

In this paper, we presented a method called UBP based on relationship strength according to the characteristics of social networks and improved the prediction accuracy of existing link prediction algorithms based on this mechanism. The method uses both the topology structure and information content in the social network. Unlike the traditional community detection algorithm experiencing issues such as randomization of community center user selection and data sparsity of user's interest, we proposed a method based on the LPA algorithm, named MSLPA. We optimized the expansion of community structure and reduced the redundancy in the community. Extensive experimental analysis on real-world datasets showed that our UBP method performs considerably better than the existing state-of-the-art methods.

As a future work, we plan to consider more real-world networks. The UBP and MSLPA methods will be evaluated on the social network for event topic detection and propagation. The UBP and MSLPA methods will be also deployed to dynamically discover and self-configure the hot events in a dynamic social network environment. The proposed algorithms will also be implemented for various different types of networks to address the existing problems in different domains.

Data Availability

The data used to support the findings of this study have not been made available because some other papers will also use this data, which is not published yet.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work reported in this paper has been supported by the National Natural Science Foundation of China Program (61502209 and 61502207) and the Suqian Municipal Science and Technology Plan Project in 2020 (S202015).

References

[1] H. Lu, S. Liu, H. Wei, and J. Tu, "Multi-kernel fuzzy clustering based on auto-encoder for fMRI functional network," *Expert Systems with Applications*, vol. 159, article 113513, 2020.

[2] A. Monney, Y. Zhan, J. Zhen, and B.-B. Benuwa, "A multi-kernel method of measuring adaptive similarity for spectral clustering," *Expert Systems with Applications*, vol. 159, pp. 1135–1147, 2020.

[3] S. Tang, S. Yuan, and Y. Zhu, "Data preprocessing techniques in convolutional neural network based on fault diagnosis towards rotating machinery," *IEEE Access*, vol. 8, pp. 149487–149496, 2020.

[4] L. L. Shi, L. Liu, Y. Wu, L. Jiang, J. Panneerselvam, and R. Crole, "A social sensing model for event detection and user influence discovering in social media data streams," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 1, pp. 141–150, 2020.

[5] S. Y. Shih, M. Lee, and C. C. Chen, "An effective friend recommendation method using learning to rank and social influence," PACIS, 2015.

[6] H. Peng, J. Li, S. Wang et al., "Hierarchical taxonomy-aware and attentional graph capsule RCNNs for large-scale multi-label text classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2505–2519, 2021.

[7] H. Peng, R. Yang, Z. Wang, J. Li, and R. Ranjan, "LIME: low-cost incremental learning for dynamic heterogeneous information networks," *IEEE Transactions on Computers*, vol. 99, p. 1, 2021.

[8] S. C. Yang, L. L. Shi, and L. Liu, "Community detection method based on user influence probability and similarity," in *2018 Sixth International Conference on Advanced Cloud and Big Data (CBD)*, pp. 183–190, Lanzhou, 2018.

[9] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[10] F. Zhang, T. Zhang, Q. Mao, and C. Xu, "A unified deep model for joint facial expression recognition, face synthesis, and face alignment," *IEEE Transactions on Image Processing*, vol. 29, pp. 6574–6589, 2020.

[11] S. Tang, S. Yuan, and Y. Zhu, "Convolutional neural network in intelligent fault diagnosis toward rotatory machinery," *IEEE Access*, vol. 8, pp. 86510–86519, 2020.

[12] J. Ge, L. -L. Shi, Y. Wu, and J. Liu, "Human-driven dynamic community influence maximization in social media data streams," *IEEE Access*, vol. 8, pp. 162238–162251, 2020.

[13] G. Salton and C. S. Yang, "On the specification of term values in automatic indexing," *Journal of Documentation*, vol. 29, no. 4, pp. 351–372, 1973.

[14] Z. Li, X. Fang, and O. R. L. Sheng, "A survey of link recommendation for social networks: methods, theoretical foundations, and future research directions," *ACM Transactions on Management Information Systems (TMIS)*, vol. 9, no. 1, pp. 1–26, 2017.

[15] M. Slokom and R. Ayachi, "A new social recommender system based on link prediction across heterogeneous networks," in *International Conference on Intelligent Decision Technologies*, pp. 330–340, Springer, Cham, 2017.

[16] S. Khusro, Z. Ali, and I. Ullah, "Recommender systems: issues, challenges, and research opportunities," in *Information Science and Applications (ICISA) 2016*, pp. 1179–1189, Springer, Singapore, 2016.

[17] T. Ha and S. Lee, "Item-network-based collaborative filtering: a personalized recommendation method based on a user's item network," *Information Processing & Management*, vol. 53, no. 5, pp. 1171–1184, 2017.

[18] S. Gregory, "Finding overlapping communities in networks by label propagation," *New Journal of Physics*, vol. 12, no. 10, p. 103018, 2010.

[19] P. Kim and S. Kim, "Detecting overlapping and hierarchical communities in complex network using interaction-based edge clustering," *Physica A: Statistical Mechanics and its Applications*, vol. 417, pp. 46–56, 2015.

[20] C. Jia, J. Ma, Q. Liu, Y. Zhang, and H. Han, "Linkboost: a link prediction algorithm to solve the problem of network vulnerability in cases involving incomplete information," *Complexity*, vol. 2020, Article ID 7348281, 14 pages, 2020.

[21] B. K. Nagra, B. Chhabra, and D. Sharma, *"Recommendation and Interest of Users", Intelligent Communication, Control and Devices*, Springer, Singapore, 2018.

[22] Y. P. Xiao, "3-HBP: a three-level hidden Bayesian link prediction model in social networks," *IEEE Transactions on Computational Social Systems*, vol. 5, pp. 430–443, 2018.

[23] Y. Li, P. Luo, Z. P. Fan, K. Chen, and J. Liu, "A utility-based link prediction method in social networks," *European Journal of Operational Research*, vol. 260, no. 2, pp. 693–705, 2017.

[24] A. K. Gupta and N. Sardana, "Naïve Bayes approach for predicting missing links in ego networks," in *2016 IEEE international symposium on Nanoelectronic and information systems (iNIS)*, pp. 161–165, Gwalior, India, 2016.

[25] V. Srinivas and P. Mitra, "Link prediction using thresholding nodes based on their degree," in *Link Prediction in Social Networks*, pp. 15–25, Springer, Cham, 2016.

[26] W. Kai, S. Liu, H. Chen, and X. Li, "A new link prediction method for complex networks based on resources carrying capacity between nodes," *Journal of Electronics & Information Technology*, vol. 41, pp. 1225–1234, 2019.

[27] E. Bastami, A. Mahabadi, and E. Taghizadeh, "A gravitation-based link prediction approach in social networks," *Swarm and Evolutionary Computation*, vol. 44, pp. 176–186, 2019.

[28] Y. Yang, J. Zhang, X. Zhu, and L. Tian, "Link prediction via significant influence," *Physica A: Statal Mechanics and its Applications*, vol. 492, pp. 1523–1530, 2018.

[29] T. S. Li, "Deep dynamic network embedding for link prediction," *IEEE Access*, vol. 6, no. 5, pp. 29219–29230, 2018.

[30] Z. X. Guo, Z. Ma, and Z. Zhang, "A novel recommendation system in location-based social networks using distributed ELM," *Memetic Computing*, vol. 10, no. 3, pp. 321–331, 2018.

[31] Z. L. Liao, L. Liu, and Y. Chen, "A novel link prediction method for opportunistic networks based on random walk and a deep belief network," *IEEE Access*, vol. 8, no. 5, pp. 16236–16247, 2020.

[32] L. Yao, L. Wang, L. Pan, and K. Yao, "Link prediction based on common-neighbors for dynamic social network," *Procedia Computer Science*, vol. 83, pp. 82–89, 2016.

[33] V. Martínez, F. Berzal, and J. C. Cubero, "A survey of link prediction in complex networks," *ACM Computing Surveys (CSUR)*, vol. 49, no. 4, pp. 1–33, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5589872, 11 pages
https://doi.org/10.1155/2021/5589872

Research Article

Attribute-Associated Neuron Modeling and Missing Value Imputation for Incomplete Data

Xiaochen Lai,[1,2] Jinchong Zhu,[1] Liyong Zhang,[3,4] Zheng Zhang,[5] and Wei Lu[3,4]

[1]*School of Software, Dalian University of Technology, Dalian 116620, China*
[2]*Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian 116620, China*
[3]*School of Control Science and Engineering, Dalian University of Technology, Dalian 116624, China*
[4]*Professional Technology Innovation Center of Distributed Control for Industrial Equipment of Liaoning Province, Dalian 116024, China*
[5]*International School of Information Science & Engineering, Dalian University of Technology, Dalian 116620, China*

Correspondence should be addressed to Liyong Zhang; zhly@dlut.edu.cn

Received 11 January 2021; Revised 9 March 2021; Accepted 9 April 2021; Published 29 April 2021

Academic Editor: Nawab Muhammad Faseeh Qureshi

The imputation of missing values is an important research content in incomplete data analysis. Based on the auto associative neural network (AANN), this paper conducts regression modeling for incomplete data and imputes missing values. Since the AANN can estimate missing values in multiple missingness patterns efficiently, we introduce incomplete records into the modeling process and propose an attribute cross fitting model (ACFM) based on AANN. ACFM reconstructs the path of data transmission between output and input neurons and optimizes the model parameters by training errors of existing data, thereby improving its own ability to fit relations between attributes of incomplete data. Besides, for the problem of incomplete model input, this paper proposes a model training scheme, which sets missing values as variables and makes missing value variables update with model parameters iteratively. The method of local learning and global approximation increases the precision of model fitting and the imputation accuracy of missing values. Finally, experiments based on several datasets verify the effectiveness of the proposed method.

1. Introduction

The interference of various factors in process of data collection, transmission and storage, etc. may cause data loss in different degrees. The incompleteness of data that leads to most of computational intelligence technologies cannot be applied directly [1]. In the cases where incomplete records cannot be simply deleted, an effective method is needed to impute missing values.

At present, researchers have proposed a variety of imputation methods. Mean imputation method imputes corresponding missing values with mean values of existing attributes [2]. The hot deck method finds the record most similar to the incomplete record in database and then imputes data with values of this record [3]. The K-nearest neighbors (KNN) imputation method takes the weighted average of K records closest to the incomplete record to impute missing values [4]. Additionally, model-based methods are usually an effective way to improve the accuracy of imputation. For example, the expectation-maximization (EM) method alternately performs the expectation step and the maximization step and iteratively updates model parameters and missing values until convergence [5]. The multiple imputation method obtains m values through one or more models and comprehensively processes the m results to impute missing values [6]. The imputation method based on linear model imputes missing values by modeling the linear relation between attributes [7]. It assumes that there is a linear correlation of the data, but the relation is complex and unknown in real data and often reflects nonlinear features.

The neural network is flexible in construction. In theory, a neural network with nonlinear activation function can approximate complex nonlinear relations [8]. The imputation

model based on neural networks can mine complex association relations within attributes of incomplete data. The imputation method based on the neural network usually uses complete records to train the network, then inputs prefilling incomplete records into the network and uses the output of network to impute missing values [9]. Sharpe and Solly [10] constructed a multilayer perceptron (MLP) for each missingness pattern, which is used to fit the regression relation between missing attributes and existing attributes. However, the number of constructed models is large, and the training is more time-consuming in the case of multiple missingness patterns. Ankaiah and Ravi [11] proposed an improved MLP imputation method, which takes each missing attribute as output and the remaining attributes as input to construct a network of single objective predictive. The number of models constructed by this method is equal to the number of missing attributes. Although the MLP imputation model can fit the regression relation between data attributes, it comes at the expense of model training time.

The auto associative neural network (AANN) is a type of network with the same number of nodes in the output layer and input layer. It is only necessary to build one model to impute incomplete data in all missingness patterns [12]. Marwala et al. [13] proposed an imputation method combining AANN and genetic algorithm (GA) and then applied it to two real datasets [14, 15]. This method takes the cost function of AANN as the fitness function of the genetic algorithm and uses the genetic algorithm to impute missing values. Based on the framework proposed by Marwala, Nelwamondo et al. adopted principal component analysis to select a reasonable number of nodes in the hidden layer [16] and reduce the dimension of data [17]. Ssali and Marwala [18] used the interval of continuous attribute divided by decision tree as data boundary, which further improved the imputation accuracy. In addition to the combination of AANN and GA, Ravi and Krishna [19] proposed four improved imputation models based on AANN, which are general regression auto associative neural network (GRAANN), particle swarm optimization based auto associative neural network (PSOAANN), auto associative wavelet neural network (AAWNN), and radial basis function auto associative neural network (RBFAANN). Among these models, GRAANN performs better than MLP and other three models in most datasets, and only needs one iteration to impute missing values. Gautam and Ravi proposed two imputation models based on AANN, which are auto associative extreme learning machine [20] and counter propagation auto associative neural network [21]. The experimental results show that the combination of local learning and global approximation can get better imputation results.

The above method only takes complete records to train the model, which avoids the problem of missing values during training. However, missing values in incomplete records will lead to incomplete model input in the imputation stage. Since the MLP imputation method constructs a specific model for each missingness pattern by taking incomplete attributes as output and complete attributes as input, it can directly input each incomplete record into the subnet of the corresponding missingness pattern. However, the AANN imputation method usually needs a prefilling method to deal with missing values during imputation. For instance, Ravi and Krishna [19] used averages to prefill missing values. Nishanth and Ravi [22] adopted K-means and K-medoids methods to prefill missing values. Gautam and Ravi [21] used the nearest neighbor method based on grey distance to prefill missing values.

The quantity of complete records is small when the missing rate in dataset is high. If only complete records are used to train the network, a large amount of information in incomplete records will be lost, and fewer records sometimes make the model unbuildable. Therefore, Silva-Ramírez et al. [23] prefilled missing values with a fixed value and then trained the network by all records. García-Laencina et al. [24–28] proposed a multitask network that uses zero to initialize missing values and allows incomplete records to participate in model training. Although the method of prefilling incomplete records with fixed values can make them participate in model training, the prefilling values have an estimation error. If the model is trained directly with prefilling data, the accuracy of the final model will be affected by the estimation error. In addition, Yoon et al. [29] proposed an imputation method base on Generative Adversarial Nets to generate data with generator. The network architecture can also try to use the inception architecture [30] in edge computing [31].

As mentioned above, the imputation method based on AANN can improve the training efficiency compared with MLP while solving multiple missingness patterns. Consequently, this paper conducts regression modeling for the attributes of incomplete data based on AANN architecture. By redesigning the data transmission structure of AANN, the representation of regression relations between data attributes is enhanced. Moreover, aiming at the problem of incomplete model input, this paper proposes a model training scheme that takes missing values as variables and makes the missing value variables update iteratively along with model parameters during model training. The improved model and training scheme make full use of the existing data in incomplete dataset and reduce the estimation error of missing value variables gradually during model training and increases the accuracy of imputation through local learning and global approximation.

The rest of this paper is organized as follows. Section 2 introduces MLP and AANN imputation models. Section 3 proposes ACFM based on AANN and a model training scheme named UMVDT. Section 4 analyses the imputation performance of ACFM and UMVDT. And the full text is summarized in Section 5.

2. MLP and AANN Imputation Models

MLP is a feed forward artificial neural network composed of input layer, output layer, and several hidden layers. When applying the MLP method to impute missing values, an MLP imputation network needs to be constructed for each missingness pattern. Figure 1 is an incomplete dataset with several missingness patterns, different positions and the number of missing values in the sample, and different deletion modes. For an incomplete data set that is missing at

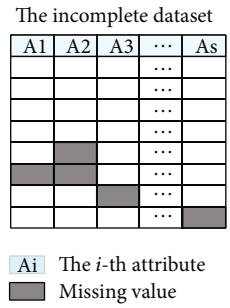


Figure 1: The incomplete dataset with multiple missing patterns.

random, the higher the missing rate, the more missing patterns. And its imputation networks are shown in Figure 2, where $x_i = [x_{i1}, x_{i2}, \cdots, x_{\text{is}}]^T$ represents the i-th record, $y_i = [y_{i1}, y_{i2}, \cdots, y_{\text{is}}]^T$ represents the network output for the i-th record, and s represents the dimension of attribute. If p_t represents indices of missing attributes in the t-th missingness pattern, the cost function of the model is

$$E_k = \frac{1}{2} \sum_{\boldsymbol{x}_i \in \boldsymbol{X}_C} \sum_{j \in P_t} \left(f_{ij} \left(\sum_{k \notin p_t} w_{jk} \cdot x_{ik} \right) - x_{ij} \right)^2, \quad (1)$$

where X_C represents the complete records, $f_{ij}(\cdot)$ represents the nonlinear mapping of the model, and w_{jk} represents the weight of the model.

AANN requires that the number of nodes in the output layer is equal to that in the input layer. In order to prevent model overfitting, the number of nodes in the hidden layer is usually set to be less than that in the input layer. As shown in Figure 3, the imputation method based on AANN can

fill incomplete data under all missingness patterns through one structure. Generally, the model is trained by complete record subset, and the incomplete record subset is reconstructed after prefilled to impute the corresponding missing values. The cost function can be expressed as

$$\mathrm{E} = \frac{1}{2} \sum_{\boldsymbol{x}_i \in \boldsymbol{X}_C} \sum_{j=1}^{s} \left(f_{ij} \left(\sum_{k=1}^{s} w_{jk} \cdot x_{ik} \right) - x_{ij} \right)^2. \quad (2)$$

It can be seen that each output value of AANN model is calculated by all input values. The output value is easier to learn the input value in the same position with model training, thus the quality of imputation values depends on a degree of the quality of pre-filling values in imputation stage. The output value of the MLP model is calculated by a regression network; so, AANN lacks clear regression relations to guide the model training and impute the missing value compared with the MLP model.

3. Proposed Architecture

3.1. Attribute CrossFitting Model. The AANN imputation model implements the imputation of multiple missingness patterns through one architecture, but it does not establish a clear regression relation between data attributes. In this paper, the regression relations between each attribute and rest attributes in incomplete dataset are expressed on one architecture by redesigning the cost function of the model

$$\mathrm{E} = \frac{1}{2} \sum_{\boldsymbol{x}_i \in \boldsymbol{X}} \sum_{j=1}^{s} \left(f_{ij} \left(\sum_{k=1, k \neq j}^{s} w_{jk} \cdot x_{ik} \right) - x_{ij} \right)^2, \quad (3)$$

where $\boldsymbol{x}$ represents an incomplete dataset. It can be seen from equation (3) that the j-th output value of the model is calculated from other input values except the j-th input value, which helps to establish a regression relation between each output value and remaining input values. Moreover, the output of the model is no longer dependent on the corresponding input value; thus, the effect of prefilling values is weakened during the imputation stage. In order to minimize the cost function, the network needs to fully learn the correlation between each output neuron and noncorresponding input neurons. Therefore, the cost function can effectively enhance the ability of mining internal association of attributes.

If the neural network is trained by incomplete records, the missing values need to be prefilled. However, there is an estimation error in prefilled values compared with original data. The model should limit the training error between prefilled data and its predicted data to optimize model parameters. This paper defines this error as missing value error. Hence, when training the network with an incomplete dataset, the cost function that the model needs to be optimized should be

$$\mathrm{E} = \frac{1}{2} \sum_{\boldsymbol{x}_i \in \boldsymbol{X}} \sum_{j \notin M_i} \left(f_{ij} \left(\sum_{k=1, k \neq j}^{s} w_{jk} \cdot x_{ik} \right) - x_{ij} \right)^2, \quad (4)$$

where M_i is the set of indexes for missing values in record x_i, and $j \notin M_i$ indicates that the missing value error is no longer used to optimize model parameters. The model constructed based on this cost function can fit regression relation between data attributes by one architecture, which is called attribute cross fitting model (ACFM) in this paper.

The data transmission process of output neurons of ACFM is shown in Figure 4 [32]. There is an incomplete record with two missing values $\hat{x}_{i1}$ and $\hat{x}_{i2}$ that input into ACFM. Because ACFM does not use missing value error to optimize model parameters, the output values y_{i1} and y_{i2} will not be calculated. The output value y_{i3} of ACFM is calculated by input values except x_{i3}. At the same time, the calculation of output values y_{i4} to y_{is} has similar processes. It can be seen that the calculation amount of ACFM is the same as that of AANN. In this article, the input data has been expanded by dimensionality times when it is implemented by programming. Then, perform forward calculation in a fully connected manner. Finally, the output is sliced, and the required value is taken out. Therefore, the parameter of ACFM in the experiment is the parameter of AANN multiplied by the number of attributes of the data.

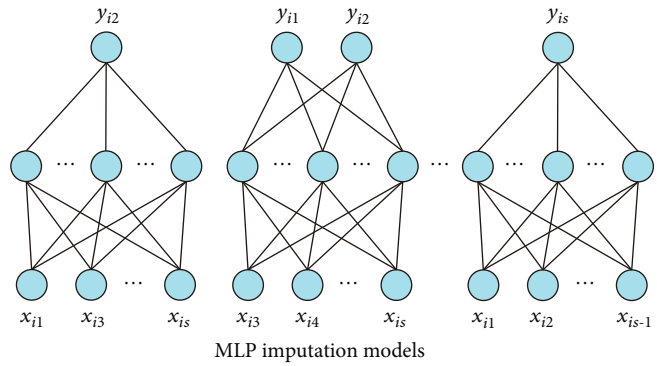


Figure 2: MLP imputation networks corresponding to each missing pattern.

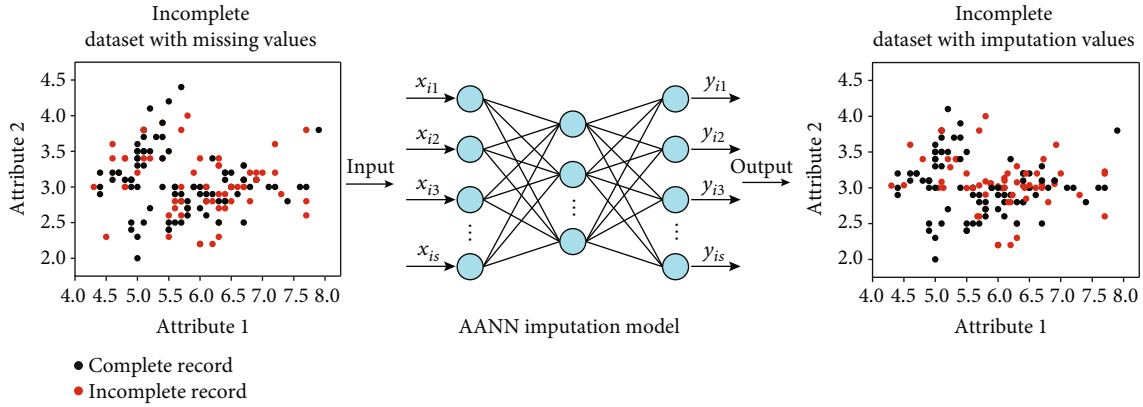


Figure 3: The diagram of AANN imputation.

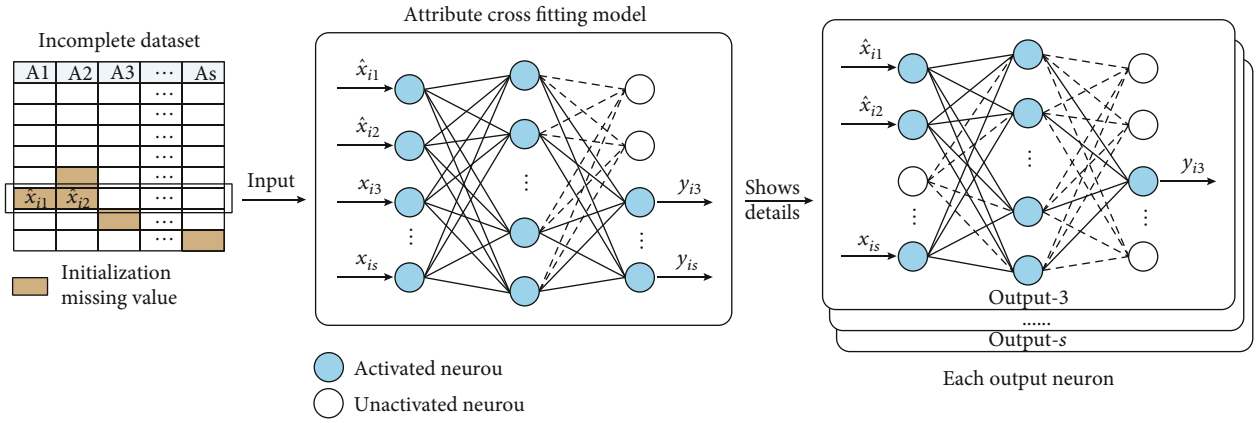


Figure 4: Schematic diagram of data transmission of ACFM, reproduced from Jinchong Zhu et al. 2020.

3.2. Updating Missing Values during Training. The prefilling missing values solve the problem of the incomplete model input, but the quality of prefilling values has an important impact on the quality of trained model when prefilling incomplete records are used to train the model directly. The prefilling values have an initial estimation error, which will reduce the accuracy of the model. Therefore, this paper proposes a model training scheme by treating missing values as variables and iteratively updating missing values during training process (UMVDT). UMVDT dynamically adjusts the values of missing and gradually reducing the estimation error of missing values, thus the missing values will meet the fitting relationship determined by existing data. As shown in Figure 5, UMVDT training scheme initializes the missing value variables in incomplete records and inputs incomplete records into ACFM for calculating the error $[e_{i1}, e_{i3},\cdots,e_{is}]$ between output and input values; then, it updates the missing value variables and the network parameters through the back propagation algorithm. The above process is repeated for all records until the model convergence. In

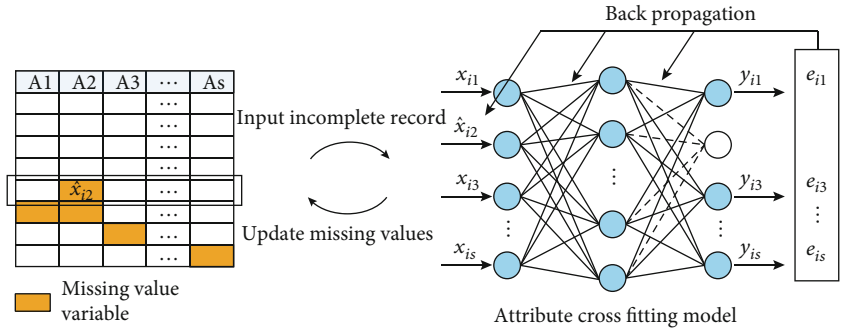


FIGURE 5: Schematic diagram of UMVDT training scheme. Reproduced from Jinchong Zhu et al. 2020.

the model based on UMVDT, the missing value variable is optimized by the regression structure within the incomplete data. The accuracy of the model will be improved with the deepening of the training, and the missing value predicted by the model will also be more accurate.

If the neurons in the input layer of ACFM are the first layer, and the output layer is $n+1$, w^l and b^l represent the weights and thresholds from layers l to $l+1$ $(1 \le l \le n)$. And each output neuron of the model is directly output after linear summation; so, it can be expressed as

$$y_{ij} = z_{ij}^{n+1} = b_j^{n+1} + \sum_{k=1}^{s_n} w_{jk}^n \cdot a_k^n, \tag{5}$$

where z_{ij}^{n+1} represents the linear summation of the j-th neuron in the layer $n+1$, s_n represents the number of neurons in layer n, and a_k^n represents the k-th output in layer n. Corresponding to each neuron j' in the output layer, the output of j-th neuron in each hidden layer can be expressed as

$$\begin{cases} \mathrm{a}_{ij}^l = g\left(z_{ij}^l\right) = g\left(b_j^{l-1} + \sum_{k=1}^{s_{l-1}} w_{jk}^{l-1} \cdot a_k^{l-1}\right), & 2 < l \le n \\ \mathrm{a}_{ij}^l = g\left(z_{ij}^l\right) = g\left(b_j^{l-1} + \sum_{k=1,k\neq j}^{'s_{l-1}} w_{jk}^{l-1} \cdot x_{ik}\right), & l = 2 \end{cases}, \tag{6}$$

where $g(\cdot)$ is the activation function. According to equation (4), the error between i -th record $\boldsymbol{x_i}$ and output $\boldsymbol{y_i}$ of the network is

$$\boldsymbol{e_i} = \frac{1}{2} \sum_{j=1,j\notin M_i}^{s_{n+1}} \left(y_{ij} - x_{ij}\right)^2. \tag{7}$$

If we define the intermediate variables δ_{ij}^{n+1} as

$$\begin{cases} \delta_{ij}^{n+1} = \dfrac{\partial e_i}{\partial z_{ij}^{n+1}} = \dfrac{\partial e_{ij}}{\partial z_{ij}^{n+1}} = \left(y_{ij} - x_{ij}\right), & \mathrm{j} \notin \mathrm{M_i} \\ \delta_{ij}^{n+1} = 0, & \mathrm{j} \in \mathrm{M_i} \end{cases} \tag{8}$$

where $j \notin M_i$ represents that the input value corresponding to the j-th predicted value is available, $j \in M_i$ represents that the input value corresponding to the j-th predicted value is missing, and thus the partial derivative is set to zero, and the corresponding model parameters are not optimized. When $2 \le l \le n$, δ_{ij}^l is

$$\delta_{ij}^l = \frac{\partial e_{ij}}{\partial z_{ij}^l} = \sum_{k=1}^{s_{l+1}} \delta_{ik}^{l+1} \cdot w_{jk}^l \cdot g'\left(z_{ij}^l\right), \tag{9}$$

and it can be concluded that the partial derivative of error e_i for the network parameter w_{jk}^l is

$$\frac{\partial \boldsymbol{e_i}}{\partial w_{jk}^l} = \frac{\partial \boldsymbol{e_i}}{\partial z_{ij}^{l+1}} \cdot \frac{\partial z_{ij}^{l+1}}{\partial w_{jk}^l} = \delta_{ij}^{l+1} \cdot a_k^l. \tag{10}$$

Similarly, the partial derivative of error e_i for the network parameter b_j^l is

$$\frac{\partial \boldsymbol{e_i}}{\partial b_j^l} = \frac{\partial \boldsymbol{e_i}}{\partial z_{ij}^{l+1}} = \delta_{ij}^{l+1}. \tag{11}$$

Assuming that the learning rate is η, and when the gradient descent method is used to optimize the model, the updating rule of the model parameters is

$$\begin{cases} w_{jk}^l = w_{jk}^l - \eta \cdot \dfrac{\partial \boldsymbol{e_i}}{\partial w_{jk}^l} \\ b_j^l = b_j^l - \eta \cdot \dfrac{\partial \boldsymbol{e_i}}{\partial b_j^l} \end{cases}. \tag{12}$$

Missing value variables are updated with the model parameters during model training. It can be deduced from equation (9) that the partial derivative of error e_i for the missing value variable $\hat{x}_{ik}$ $(k \in M_i)$ is

$$\frac{\partial \boldsymbol{e_i}}{\partial \hat{x}_{ik}} = \sum_{j=1}^{s_2} \delta_{ij}^2 \cdot w_{jk}^1, \tag{13}$$

INPUT: complete dataset D, missing rate, ACFM, learning rate η, maximum rounds T.
OUTPUT: the imputation error of D at specified missing rate.
Generate an incomplete dataset D' according to specified missing rate.
Initialize missing values as variables, model weights, and thresholds.
Set t =0, precision =1.
while t<T and precision<0.001 do.
  $t = t + 1$.
  for x in D':
    Input x into model and get output y.
    Calculate the error for updating the model parameters and missing value variables respectively.
  end for
  Reconstruct model output and predict missing values.
  Calculate the imputation error and precision.
end while
Output the imputation error.

Algorithm 1: The imputation based on ACFM and UMVDT.

and the updating rule of missing value variable $\widehat{x}_{ik}$ is

$$\widehat{x}_{ik} = \widehat{x}_{ik} - \eta \cdot \frac{\partial \boldsymbol{e_i}}{\partial \widehat{x}_{ik}}. \tag{14}$$

In summary, the imputation algorithm based on the ACFM model and UMVDT training scheme is described as follows:

Table 1: Description of datasets.

Datasets	Records	Attributes	Datasets	Records	Attributes
Blood	748	4	Iris	150	4
Buddymove	249	6	Seeds	210	7
Ecoli	336	7	Stock	252	12
Glass	214	10	Wine	178	13
Concrete	1030	9	Abalone	4177	7

4. Experiment

4.1. Datasets. In order to verify the imputation performance of proposed method, ten complete datasets obtained from the UCI database are used in our experiment, and the description of datasets is shown in Table 1. Among them, Stock is often used for clustering tasks, Concrete is often used for regression tasks, and the remaining data sets can be used for classification tasks. Most of these data are numeric, and some of them are nonnumeric in the ID column, which was deleted in the experiment. Additional information can refer to data sets UCI official website. For the sake of forming incomplete datasets, partial data are deleted randomly according to specified deletion rates which are set as 5%, 10%, 15%, 20%, 25%, and 30% and ensure that each incomplete record has at least one attribute value, which can be used for normal training.

4.2. Experimental Design. Six imputation methods based on MLP, AANN, and ACFM are realized. The method based on AANN and ACFM realizes the training by traditional training scheme and UMVDT training scheme. Traditional training scheme only uses the mean value to prefill missing value, and does not update missing values. To verify the effect of missing value error on imputation accuracy of the model, this paper uses equation (3) with missing value error and equation (4) without missing value error as the cost function, respectively. The specific methods are described as follows:

(1) The imputation method based on the MLP model and traditional training scheme (MLP-I): taking missing attributes as output and other attributes as input, multiple networks of single objective predictive are established based on MLP. These models are trained with complete records during the training stage. In the imputation stage, the incomplete records are prefilling with the mean method, and missing values are imputed with the reconstructed model output

(2) The imputation method based on the AANN model and traditional training scheme (AANN-I): the imputation process is same as MLP-I, but the architecture is AANN

(3) The imputation method based on the ACFM model where missing value error is used to optimize model parameters (ACFM-MEI): equation (3) is used as the cost function of ACFM. The incomplete records are prefilling with the mean method, and then all records are used to train the model. Finally, the reconstructed model output is used to impute missing values

(4) The imputation method based on the ACFM model where missing value error is not used to optimize model parameters (ACFM-I): equation (4) is used as the cost function of ACFM, and the process is same as ACFM-MEI

(5) The imputation method based on the AANN model and UMVDT training scheme (AANN-UMVDT): the mean value is used to initialize missing value variables. After that, the method uses all data to train the

Table 2: The MAPE values of first five datasets.

Datasets	Methods	Missing rates					
		5%	10%	15%	20%	25%	30%
Blood	MLP-I	1.113	1.122	0.872	0.68	0.861	0.985
	AANN-I	0.983	1.087	1.274	1.114	1.188	1.212
	AANN-UMVDT	0.941	0.968	1.005	0.974	1.026	1.124
	ACFM-MEI	0.513	0.575	0.683	0.728	0.787	0.854
	ACFM-I	0.488	0.537	0.62	0.671	0.728	0.764
	ACFM-UMVDT	0.449	0.51	0.599	0.633	0.754	0.8
Buddymove	MLP-I	0.24	0.151	0.166	0.199	0.213	0.237
	AANN-I	0.202	0.234	0.243	0.272	0.261	0.292
	AANN-UMVDT	0.14	0.147	0.159	0.163	0.167	0.177
	ACFM-MEI	0.122	0.149	0.172	0.199	0.2	0.222
	ACFM-I	0.122	0.142	0.164	0.184	0.186	0.197
	ACFM-UMVDT	0.105	0.115	0.131	0.15	0.15	0.176
Ecoli	MLP-I	0.472	0.231	0.233	0.236	0.274	0.285
	AANN-I	0.192	0.273	0.281	0.259	0.275	0.304
	AANN-UMVDT	0.185	0.247	0.261	0.238	0.257	0.276
	ACFM-MEI	0.159	0.226	0.253	0.237	0.259	0.283
	ACFM-I	0.157	0.22	0.241	0.226	0.247	0.269
	ACFM-UMVDT	0.155	0.212	0.248	0.232	0.241	0.269
Glass	MLP-I	0.287	0.351	0.298	0.343	0.423	0.45
	AANN-I	0.376	0.405	0.407	0.363	0.417	0.439
	AANN-UMVDT	0.407	0.429	0.403	0.356	0.414	0.419
	ACFM-MEI	0.311	0.342	0.344	0.303	0.357	0.354
	ACFM-I	0.29	0.346	0.338	0.314	0.35	0.35
	ACFM-UMVDT	0.286	0.327	0.346	0.314	0.353	0.345
Iris	MLP-I	0.157	0.15	0.237	0.234	0.272	0.335
	AANN-I	0.298	0.358	0.376	0.386	0.401	0.455
	AANN-UMVDT	0.15	0.158	0.19	0.188	0.189	0.234
	ACFM-MEI	0.17	0.186	0.25	0.279	0.297	0.367
	ACFM-I	0.153	0.139	0.219	0.217	0.237	0.272
	ACFM-UMVDT	0.139	0.128	0.167	0.173	0.186	0.234

model, dynamically updates missing value variables during model training, and reconstructs the model output to impute missing values

(6) The imputation method based on the ACFM model and UMVDT training scheme (ACFM-UMVDT): the process is same as AANN-UMVDT, but the architecture is ACFM

All models are optimized based on the gradient descent method with momentum. The learning rate is set as 0.2, and momentum is set as 0.9. All methods were repeated ten times at each missing rate, and the average value of ten imputation errors was taken as experimental results. Imputation error is evaluated by mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{\sum_{i=1}^{n} |M_i|} \sum_{\boldsymbol{x_i} \in X_I} \sum_{j \in M_i} \left| \frac{y_{ij} - x_{ij}}{x_{ij}} \right|, \tag{15}$$

where X_I represents incomplete records subset, and $\sum_{i=1}^{n} |M_i|$ represents the number of missing values.

4.3. Experimental Results. The experimental results are shown in Tables 2 and 3.

4.4. Experimental Discussion. The impact of architecture on imputation results: by observing the MAPE values of ACFM-I, AANN-I, and MLP-I in Tables 2 and 3, we can see that the results of ACFM-I are slightly worse than those of MLP-I in four cases, which is Ecoli at the missing rate of 15%, Glass at the missing rates of 5% and 15%, and Stock at the missing rate of 5%. In addition to the above, all the results of ACFM-I are better than MLP-I and AANN-I. Besides, there are forty-three results of MLP-I superior to the AANN-I among sixty imputation results. This result shows that MLP can more accurately characterize the regression relation within dataset than AANN, thereby obtaining higher imputation accuracy. ACFM increases the ability to

Table 3: The MAPE values of last five datasets.

Datasets	Methods	Missing rates					
		5%	10%	15%	20%	25%	30%
Seeds	MLP-I	0.071	0.093	0.104	0.114	0.096	0.151
	AANN-I	0.083	0.096	0.095	0.109	0.097	0.122
	AANN-UMVDT	0.067	0.077	0.072	0.084	0.076	0.085
	ACFM-MEI	0.07	0.08	0.08	0.097	0.088	0.099
	ACFM-I	0.067	0.077	0.076	0.09	0.083	0.09
	ACFM-UMVDT	0.062	0.068	0.067	0.081	0.076	0.088
Stock	MLP-I	0.109	0.145	0.171	0.227	0.279	0.341
	AANN-I	0.181	0.203	0.22	0.236	0.268	0.32
	AANN-UMVDT	0.177	0.185	0.185	0.194	0.19	0.201
	ACFM-MEI	0.123	0.15	0.159	0.175	0.176	0.186
	ACFM-I	0.113	0.144	0.154	0.168	0.172	0.176
	ACFM-UMVDT	0.102	0.126	0.14	0.17	0.181	0.186
Wine	MLP-I	0.183	0.215	0.282	0.324	0.372	0.488
	AANN-I	0.211	0.25	0.246	0.294	0.352	0.392
	AANN-UMVDT	0.199	0.214	0.205	0.208	0.215	0.233
	ACFM-MEI	0.172	0.197	0.198	0.204	0.211	0.225
	ACFM-I	0.176	0.185	0.189	0.196	0.201	0.21
	ACFM-UMVDT	0.175	0.186	0.194	0.199	0.206	0.215
Concrete	MLP-I	0.452	0.402	0.401	0.45	0.481	0.57
	AANN-I	0.465	0.405	0.478	0.519	0.524	0.569
	AANN-UMVDT	0.501	0.429	0.466	0.498	0.493	0.511
	ACFM-MEI	0.3	0.288	0.32	0.372	0.372	0.403
	ACFM-I	0.31	0.298	0.331	0.383	0.361	0.386
	ACFM-UMVDT	0.336	0.342	0.4	0.436	0.431	0.504
Abalone	MLP-I	0.155	0.219	0.352	0.499	0.451	0.631
	AANN-I	0.567	0.547	0.605	0.633	0.632	0.535
	AANN-UMVDT	0.133	0.114	0.154	0.189	0.191	0.337
	ACFM-MEI	0.184	0.212	0.248	0.336	0.381	0.467
	ACFM-I	0.145	0.163	0.196	0.223	0.243	0.246
	ACFM-UMVDT	0.119	0.137	0.125	0.168	0.171	0.193

fit regression relations by modifying the cost function compared to AANN. Meanwhile, compared with MLP, ACFM fits multiple regression relations through one architecture, which increases the generalization ability of ACFM on the premise of improving the imputation accuracy.

The impact of missing value error on imputation results: it can be observed from Tables 2 and 3 that ACFM-I performs slightly worse than ACFM-MEI under the missing rates of 10% and 20% in the Glass dataset, 5% in the Wine dataset, and four kinds of missing rates on the Concrete dataset. In addition, the imputation result of ACFM-I is better than ACFM-MEI in imputation results. It shows that the optimization of model parameters by missing value error affects the accuracy of modeling and thus leads to the poor performance of imputation results.

Taking the Iris dataset as an example, the imputation results of ACFM-MEI and ACFM-I at missing rates of 5%-30% are shown in Figure 6. With the increase of missing rate, the gap between imputation results of ACFM-MEI and ACFM-I becomes larger. If we continue to use the missing value error to optimize the model parameters when there are more and more missing values in the dataset, the deviation of model will also increase. Therefore, in this paper, equation (4) is used as the cost function of ACFM in this paper; that is, only the errors of existing data are used to optimize the model parameters, which have certain reasonableness and correctness.

Comparison between UMVDT and traditional training scheme: except for the results of Glass and Concrete datasets at missing rates of 5% and 10%, the results of AANN-UMVDT are better than those of AANN-I. Among the 60 imputation results, the results of ACFM-UMVDT are better than those of ACFM-I accounted for 66.7%, wherein the Concrete data set of data values vary greatly. There are many zero values and large values, and many samples have the same value in attributes. UMVDT will change the missing values during the training process, which may cause the imputation results of many samples with the same value to

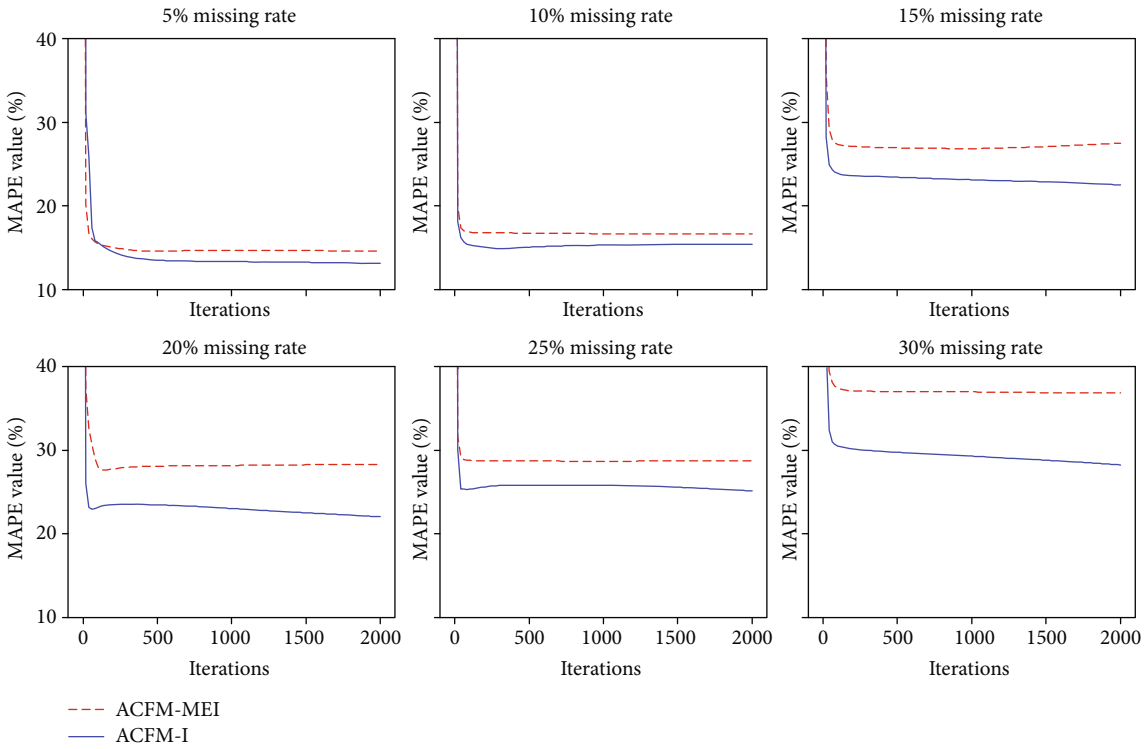


Figure 6: MAPE values of ACFM-I and ACFM-MEI on the Iris dataset.

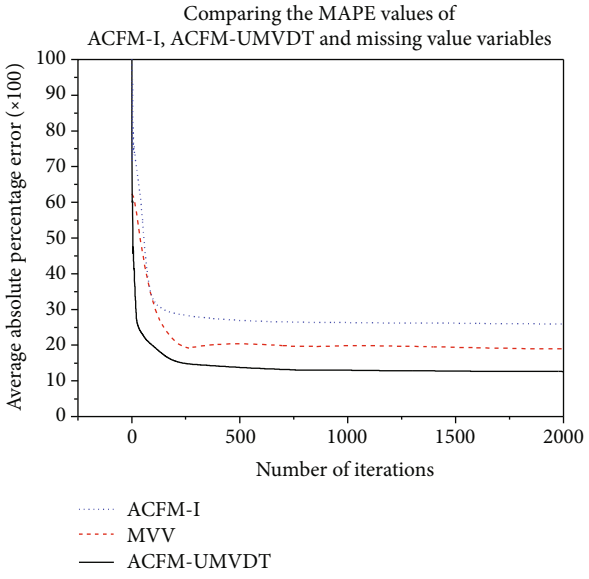


Figure 7: The imputation results of ACFM-I and ACFM-UMVDT and updating results of missing value variables on the Iris dataset, reproduced from Jinchong Zhu et al. 2020.

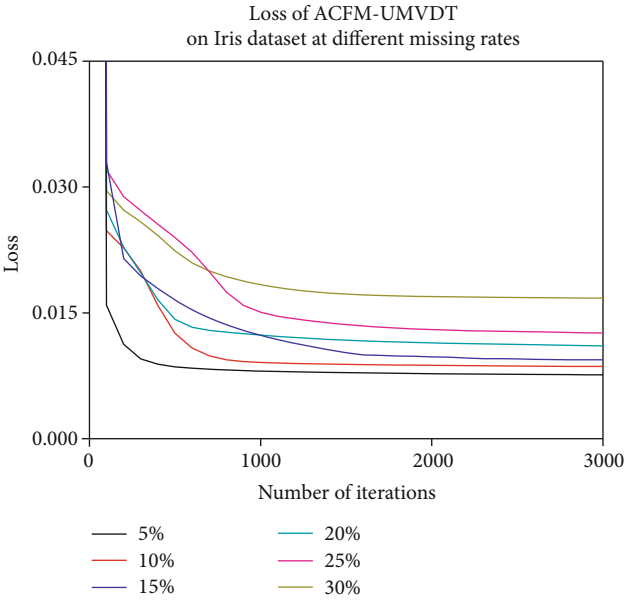


Figure 8: Convergence curve of fitting error of ACFM-UMVDT on the Iris dataset.

be unstable. The above results show that UMVDT training scheme has higher imputation performance than traditional one. UMVDT training scheme makes full use of the whole existing data in incomplete records and takes missing values as variables to make them gradually match the fitting relationship. The missing value variables and model parameters are updated alternately; so, the imputation effect can be improved significantly.

When the missing rate of the Iris dataset is 15%, the imputation of ACFM-I and ACFM-UMVDT and the variation of the missing value variables (MVV) of ACFM-UMVDT in each round are shown in Figure 7. It can be found that the missing values tend to be stable soon after a short period of fluctuation, and the imputation results of ACFM-UMVDT also tend to be stable with the increase of iteration rounds. The missing value is updated iteratively.

Not only the MAPE values calculated by missing value variables are more accurate than those of original model but also the imputation accuracy can be further improved by the model which is trained by the data updated iteratively.

The convergence of the proposed method: we take the Iris dataset as an example to verify the convergence of the proposed method. Figure 8 shows the fitting error of ACFM-UMVDT at various missing rates. It can be observed that all curves of fitting errors decrease in different degrees at beginning and become stabilized gradually. It is because the UMVDT training scheme constantly updates missing value variables and changes missing values in incomplete records. Missing value variables and model parameters converge continuously in the alternate updating process. The curves in Figure 8 show that the imputation method proposed in this paper has ideal convergence.

5. Conclusions

To solve the problem of imputation of missing values, this paper conducts attribute association modeling for incomplete data based on AANN. By modifying the cost function of AANN, this paper represents the regression relation between each attribute and the rest attributes of incomplete data on one architecture and redesign ACFM for enhanced to fit the association relation between data attributes. And we only use the training errors of existing data to optimize the model to reduce the inaccurate error between missing values and its predicted values in incomplete data to optimize the model. In addition, for the problem of incomplete model input, this paper proposes UMVDT training scheme, which sets missing values as variables and updates the model parameters and missing value variables alternately. UMVDT gradually optimizes the missing value variables through the regression structure of the model and further reduces the negative impact of the uncertainty of missing values during model input on the model. Experimental results show that the ACFM model can obtain more accurate imputation results compared with MLP and AANN models, and UMVDT further improves the accuracy of imputation on AANN and ACFM models by gradually iterating the missing value variables compared with traditional training scheme.

Data Availability

All datasets in this study can be downloaded from http://archive.ics.uci.edu/ml/datasets.php. And all experimental results are included in this published article.

Conflicts of Interest

We declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of China (62076050, 62073056) and National Key R&D Program of China (2018YFB1700200).

References

[1] F. V. Nelwamondo, D. Golding, and T. Marwala, "A dynamic programming approach to missing data estimation using neural networks," *Information Sciences*, vol. 237, pp. 49–58, 2013.

[2] S. M. C. M. Nor, S. M. Shaharudin, S. Ismail, N. H. Zainuddin, and M. L. Tan, "A comparative study of different imputation methods for daily rainfall data in east-coast peninsular Malaysia," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 2, pp. 635–643, 2020.

[3] V. R. Elgin Christo, H. Khanna Nehemiah, B. Minu, and A. Kannan, "Correlation-Based Ensemble Feature Selection Using Bioinspired Algorithms and Classification Using Backpropagation Neural Network," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 7398307, 17 pages, 2019.

[4] I. B. Aydilek and A. Arslan, "A novel hybrid approach to estimating missing values in databases using k-nearest neighbors and neural networks," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 8, pp. 4705–4717, 2012.

[5] Z. Ghahramani and M. Jordan, "Supervised learning from incomplete data via an EM approach," in *Advances in Neural Information Processing Systems*, J. Cowan, G. Tesauro, and J. Alspector, Eds., vol. 6, pp. 120–127, Morgan-Kaufmann, 1994.

[6] R. Suphanchaimat, S. Limwattananon, and W. Putthasri, "Multiple imputation technique: handling missing data in real world health care research," *Southeast Asian Journal of Tropical Medicine and Public Health*, vol. 48, no. 3, pp. 694–703, 2017.

[7] K. Yang, J. Li, and C. Wang, "Missing values estimation in microarray data with partial least squares regression," in *International Conference on Computational Science*, pp. 662–669, Springer, Berlin, Heidelberg, May 2006.

[8] A. Tealab, "Time series forecasting using artificial neural networks methodologies: a systematic review," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 334–340, 2018.

[9] I. A. Gheyas and L. S. Smith, "A neural network-based framework for the reconstruction of incomplete data sets," *Neurocomputing*, vol. 73, no. 16-18, pp. 3039–3065, 2010.

[10] P. K. Sharpe and R. J. Solly, "Dealing with missing values in neural network-based diagnostic systems," *Neural Computing & Applications*, vol. 3, no. 2, pp. 73–77, 1995.

[11] N. Ankaiah and V. Ravi, "A novel soft computing hybrid for data imputation," in *Proceedings of the International Conference on Data Mining (DMIN) (P. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp)*, Las Vegas, USA, 2011.

[12] L. Gondara and K. Wang, "Mida: multiple imputation using denoising autoencoders," in *Pacific-Asia conference on knowledge discovery and data mining*, pp. 260–272, Springer, Cham, 2018.

[13] M. Abdella and T. Marwala, "The use of genetic algorithms and neural networks to approximate missing data in database," in *IEEE 3rd international conference on computational cybernetics, 2005. ICCC 2005*, pp. 207–212, Mauritius, April 2005.

[14] B. L. Betechuoh, T. Marwala, and T. Tettey, "Autoencoder networks for HIV classification," *Current Science*, vol. 91, no. 11, pp. 1467–1473, 2006.

[15] T. Marwala and S. Chakraverty, "Fault classification in structures with incomplete measured data using autoassociative

neural networks and genetic algorithm," *Current Science*, vol. 90, no. 4, pp. 542–548, 2006.

[16] F. J. Mistry, F. V. Nelwamondo, and T. Marwala, "Missing data estimation using principle component analysis and autoassociative neural networks," *Journal of Systemics, Cybernatics and Informatics*, vol. 7, no. 3, pp. 72–79, 2009.

[17] A. K. Mohamed, F. V. Nelwamondo, and T. Marwala, "Estimating missing data using neural network techniques, principal component analysis and genetic algorithms," in *Proceedings of the Eighteenth Annual Symposium of the Pattern Recognition Association of South Africa*, Pietermaritzburg, South Africa, 2007.

[18] G. Ssali and T. Marwala, "Estimation of missing data using computational intelligence and decision trees," 2007, https://arxiv.org/abs/0709.1640.

[19] V. Ravi and M. Krishna, "A new online data imputation method based on general regression auto associative neural network," *Neurocomputing*, vol. 138, pp. 106–113, 2014.

[20] C. Gautam and V. Ravi, "Data imputation via evolutionary computation, clustering and a neural network," *Neurocomputing*, vol. 156, pp. 134–142, 2015.

[21] C. Gautam and V. Ravi, "Counter propagation auto-associative neural network based data imputation," *Information Sciences*, vol. 325, pp. 288–299, 2015.

[22] K. J. Nishanth and V. Ravi, "A computational intelligence based online data imputation method: an application for banking," *Journal of Information Processing Systems*, vol. 9, no. 4, pp. 633–650, 2013.

[23] E. L. Silva-Ramírez, R. Pino-Mejías, M. López-Coello, and M. D. Cubiles-de-la-Vega, "Missing value imputation on missing completely at random data using multilayer perceptrons," *Neural Networks*, vol. 24, no. 1, pp. 121–129, 2011.

[24] P. J. García-Laencina, J.-L. Sancho-Gómez, and A. R. Figueiras-Vidal, "Classifying patterns with missing values using multi-task learning perceptrons," *Expert Systems with Applications*, vol. 40, no. 4, pp. 1333–1341, 2013.

[25] P. J. García-Laencina, J. L. Sancho-Gómez, and A. R. Figueiras-Vidal, "Pattern classification with missing data: a review," *Neural Computing and Applications*, vol. 19, no. 2, pp. 263–282, 2010.

[26] J. M. Jerez, I. Molina, P. J. García-Laencina et al., "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," *Artificial Intelligence in Medicine*, vol. 50, no. 2, pp. 105–115, 2010.

[27] P. J. García-Laencina, J. Serrano, A. R. Figueiras-Vidal, and J. L. Sancho-Gómez, "Multi-task neural networks for dealing with missing inputs," in *International work-conference on the interplay between natural and artificial computation*, pp. 282–291, Springer, Berlin, Heidelberg, June 2007.

[28] P. J. García-Laencina, J. Sancho-Gomez, and A. R. Figueiras-Vidal, "Pattern classification with missing values using multi-task learning," in *The 2006 IEEE international joint conference on neural network proceedings*, pp. 3594–3601, Vancouver, BC, Canada, July 2006.

[29] J. Yoon, J. Jordon, and M. Schaar, "Gain: missing data imputation using generative adversarial nets," in *International conference on machine learning*, vol. 80, pp. 5689–5698, Stockholm, Sweden, July 2018, PMLR.

[30] X. Kong, K. Wang, S. Wang et al., "Real-time mask identification for COVID-19: an edge computing-based deep learning framework," *IEEE Internet of Things Journal*, 2021.

[31] X. Kong, S. Tong, H. Gao et al., "Mobile edge cooperation optimization for wearable internet of things: a network representation-based framework," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 5050–5058, 2021.

[32] J. Zhu, L. Zhang, X. Lai, and G. Zhang, "Imputation of incomplete data based on attribute cross fitting model and iterative missing value variables," in *International symposium on neural networks*, pp. 167–175, Springer, Cham, December 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5520455, 13 pages
https://doi.org/10.1155/2021/5520455

Research Article

Optimal Workload Allocation for Edge Computing Network Using Application Prediction

Zhenquan Qin,[1,2] **Zanping Cheng**,[1,2] **Chuan Lin**,[3] **Zhaoyi Lu**,[1,2] **and Lei Wang**[1,2]

[1]*Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, China*
[2]*School of Software, Dalian University of Technology, 116620, China*
[3]*Software College, Northeastern University, 110819, China*

Correspondence should be addressed to Chuan Lin; chuanlin1988@gmail.com

Received 8 January 2021; Revised 16 February 2021; Accepted 9 March 2021; Published 25 March 2021

Academic Editor: Varun Menon

By deploying edge servers on the network edge, mobile edge computing network strengthens the real-time processing ability near the end devices and releases the huge load pressure of the core network. Considering the limited computing or storage resources on the edge server side, the workload allocation among edge servers for each Internet of Things (IoT) application affects the response time of the application's requests. Hence, when the access devices of the edge server are deployed intensively, the workload allocation becomes a key factor affecting the quality of user experience (QoE). To solve this problem, this paper proposes an edge workload allocation scheme, which uses application prediction (AP) algorithm to minimize response delay. This problem has been proved to be a NP hard problem. First, in the application prediction model, long short-term memory (LSTM) method is proposed to predict the tasks of future access devices. Second, based on the prediction results, the edge workload allocation is divided into two subproblems to solve, which are the task assignment subproblem and the resource allocation subproblem. Using historical execution data, we can solve the problem in linear time. The simulation results show that the proposed AP algorithm can effectively reduce the response delay of the device and the average completion time of the task sequence and approach the theoretical optimal allocation results.

1. Introduction

In recent years, the popularity of mobile devices, such as smart phones or tablet computers, has a huge impact on mobile and wireless networks, which has triggered the challenges of global mobile networks. In addition, more IoT devices have begun to be promoted and used on a large scale. IoT devices and their applications have produced a lot of data and network traffic. A new report by International Data Corporation (IDC) estimates that there will be 41.6 billion IoT devices connected to the Internet, which will generate 79.4 zettabytes (ZB) of data in 2025 [1]. However, the current infrastructure uses cloud computing as the underlying computing support, which is a centralized processing model. With the trend of exponential growth of IoT devices, the transmission of terminal computing tasks and storage tasks to the cloud computing will lead to problems such as high cloud service delay, high bandwidth occupancy, and security privacy in the big data network. Faced with these problems, edge computing emerges as the supplement of cloud computing. Different from the centralized computing model of cloud computing, edge computing is a distributed computing processing model. By deploying edge servers on the network edge, mobile edge computing provides resources and services to minimize the cloud load. Meanwhile, mobile edge computing also complements cloud computing by enhancing IoT user services for delay-sensitive applications [2].

Although the edge servers can provide computing or storage resources in edge computing network [3], the computing or storage resources of edge servers are limited. Intelligent and practical algorithms are needed to schedule and allocate resources to ensure the optimal resource allocation. Thus, efficient workload allocation strategy is extremely important, which directly determines the utilization efficiency of resources. The current academic research also focuses on how to allocate the load reasonably from the

existing resources of edge servers, dynamically according to the application of access devices and network conditions [4]. However, the resource allocation for different types of applications also affects the response delay of different types of requests. Since the computing size per request for different applications are different, the computing capacity of edge server should be optimally allocated for different applications in order to reduce the response delay of all applications of user equipment.

To solve the above problem, this paper proposes a workload allocation strategy that can be applied in real scenarios to make better use of resources in edge network and reduce the response delay of edge devices. Through the analysis of the maximum information entropy, Song et al. [5] found that the network activity information of terminal equipment has 93% predictability. Therefore, in the direction of improving the network system, the analysis of mobile data has great potential. On the one hand, mobile traffic information is very important to clearly describe how mobile terminal users use access network resources. On the other hand, the analysis rules of mobile traffic information can better guide the deployment of network system and the allocation of resources. To enable future networks to achieve "no perception" of latency, this paper first proposes AP algorithm to minimize the average task completion time, which takes into account transmission delay and computing delay. The main contributions of this paper are as follows:

(i) We innovatively put forward the concept of application prediction in edge computing architecture to solve edge workload allocation problem

(ii) In the application prediction model, the AP algorithm based on LSTM uses historical data to predict the possible tasks of future access device. By application prediction, the virtual machine is loaded on the edge server in advance, which reduces the service response delay. F-Measure parameter is weighted harmonic average of precision and recall, which is introduced to evaluate the accuracy of AP algorithm

(iii) Based on the prediction results of AP algorithm, workload allocation problem is divided into two subproblems, which are the task assignment problem and the resource allocation subproblem within the edge server

The rest of the paper is organized as follows. Section 2 introduces the related work. Section 3 describes the system, transmission delay, and computing delay models and then formulates the problem. Section 4 describes the detailed algorithm for solving workload allocation problem. Simulation studies are conducted to demonstrate the performance of the proposed algorithm in Section 5. Section 6 concludes this paper.

2. Related Works

On the edge server side, the edge workload allocation scheme refers to how to best allocate existing resources to complete the current task when task offloading by different devices forms a task queue.

Many existing work studies computing offloading decision and involved resource allocation to optimize the time delay or energy consumption of devices in edge scenarios. Mao et al. in [6] and Lyu et al. in [7] designed the task load allocation algorithm based on the Lyapunov optimized framework and implemented the offloading decision only based on the current system state. Chen et al. in [8] proposed the joint optimization of offloading decision and allocation of communication and computing resources. Semiqualitative relaxation and a novel random mapping method were adopted to effectively solve the NP-hard problem, so as to minimize the total system cost. Du et al. in [9] designed a fairness guarantee algorithm for computing offloading and resource allocation, which jointly optimized the offloading decision and the allocation of computing resources, transmission power, and radio bandwidth, while ensuring user fairness and maximum tolerable delay. Liu et al. in [10] conducted an in-depth study on the energy consumption, payment cost, and delay of the offloading process in the edge computing system by using queuing theory. Tan et al. in [11] proposed an online task assignment and scheduling algorithm, adding a weight to the response time of each task to represent its delay sensitivity, thus assigning greater weight to delay-sensitive tasks so as to provide them with higher priority. Reference [12] proposed a new mobility management scheme in the ultradense MEC network, using Lyapunov optimization and the theory of the dobby bandit to optimize the delay of the equipment under the constraint of long-term energy consumption.

Some of the existing work is also studying the horizontal collaboration between edge servers. Jia et al. in [13] pointed out that cooperation between edge servers can balance computing load and generate huge performance improvements, which is far more than the situation of only performing tasks on edge servers alone. Xiao and Krunz in [14] proposed a collaborative task forwarding strategy to improve the user's QoE. Fan and Ansari in [15] designed a task type based allocation scheme to minimize response time by determining the computing destination and allocated resources for each task.

In 2019's section, many new strategies for reducing delay in resource allocation are also proposed. In [16], Dlamini and Gambin proposed an adaptive fog architecture, which dynamically selects different links to reduce transmission delay by measuring link delay. When the link delay is high, the performance is poor. Acharya in [17] proposed that for specific computer vision programs, services that may be needed should be loaded in advance during the offline phase of the program, so as to reduce the delay of service loading when the program is running. However, if the application is changed or too many applications are opened at the same time, the performance of this strategy will decline sharply. Xu et al. in [18] applied the technology of service caching. When the application is changed frequently, the algorithm has poor effect on the delay reduction. Shu et al. in [19] used the topology diagram and unique sequence diagram of tasks for hierarchical offloading and resource loading, but the fine-grained topological ordering of tasks needed to be

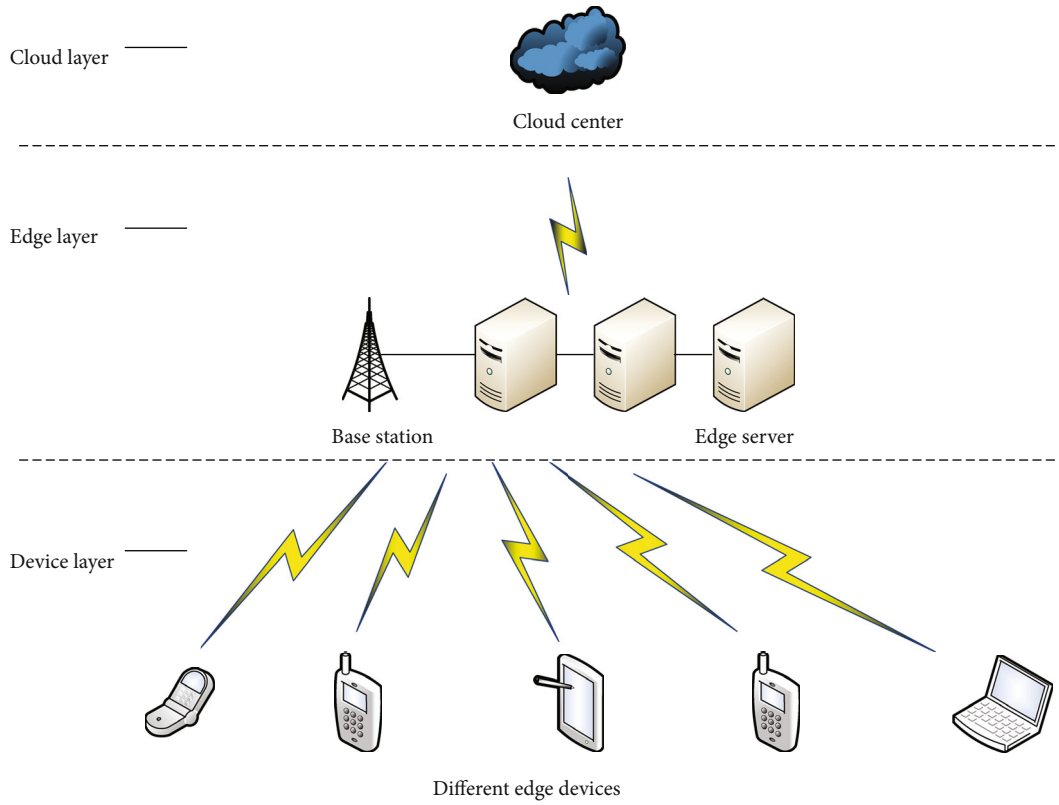


Figure 1: Edge computing workload architecture model.

determined in advance, which was not suitable for the changeable network environment.

In summary, the research boundary of edge computing or fog computing is fuzzy, but it is unified in the optimization of task assignment decision and resource coordination allocation in [20]. However, all existing research schemes have some limitations and ignore the importance of historical data. In this paper, an innovative concept based on application prediction is proposed, using historical data to predict the application to be used in the future in the edge computing network architecture. By loading the virtual service of the application on the edge server in advance, the service loading delay can be minimized, thus reducing the overall edge task response time.

3. System Model and Problem Formulation

This section briefly introduces the system, transmission delay, and computing delay models as well as problem formulation.

3.1. System Model. Figure 1 shows the three-layer edge computing workload allocation architecture adopted.

The device layer contains different edge devices, each of which commits the computing task at different times. J represents the set of edge devices (EDs). j represents different device. K represents the set of applications (APPs), all of which are in a known collection. k represents different types of application. T represents the network time slot. T_n represents the current time slot at time n. T_{n-1} represents the previous network time slot. T_{n+1} represents the next network slot. In each network slot, the tasks that device j submits to the edge layer can be represented by K_j.

In the edge layer, BS is used to represent the base station, and the User App LCM Proxy module is set up, which is responsible for receiving the task sequence from different mobile edge devices and uniformly assigning the tasks to different servers. The computing unit is different edge server hosts, with M representing the set of edge servers (ESs) and m representing different edge servers. In each edge server, B_m represents total bandwidth of edge server m, B_m^{idle} represents the remaining allocable bandwidth, C_m represents total computing resources, and C_m^{idle} represents the remaining allocable computing resources in CPU cycles per second.

In workload allocation scheme, transmission delay and computing delay are taken into account. Hence, the task requests of edge device are more likely to be allocated to the closer or lighter workload edge server. The meaning of the parameters used in this paper is listed in Table 1.

3.2. Transmission Delay Model. When the edge device task request is assigned to a different edge server, the edge device will establish a connection with the edge server. The data generated in the edge device will be transferred to the edge server. Therefore, transmission delay consists of two parts: link delay and data delay.

In general, link delay is related to channel state and power, which is not discussed in this paper. t_{jm}^{link} is used to represent transmission link delay between device j and edge

Table 1: The representative meaning of different symbols.

Symbol	Description
J	The set of EDs
K	The set of APPs
T	The different network time slots
M	The set of ESs
K_j	The task that edge device j commits to the edge layer
B_m	The total bandwidth of edge server m
B_m^{idle}	The remaining allocable bandwidth
B_{jm}	The bandwidth allocated on edge server m for the task of device j
C_m	The total computing resources of edge server m
C_m^{idle}	The remaining allocable computing resources
C_{jm}	The computing resources allocated on edge server m for the task of device
t_{jm}^{link}	The transmission link delay between device j and edge server m
t_{jm}^{tran}	The total transmission delay between device j and edge server
x_{jm}	In binary array, whether the task of device j is assigned to the edge server m
t_k^{load}	The APP k service load time on the edge server side
θ_{jk}	The calculation strength required for task k unit data of device
D_{jk}	The amount of data to be processed by task k of device j
τ_{jk}^{worse}	The maximum tolerance delay by task k of device j

server m, and the multiaccess edge coordinator in the edge computing system is responsible for update and maintenance.

In each network slot, the task that device j submits to the edge layer can be represented by K_j. In task K_j, the data size to be transmitted and processed is D_{jk}. After the task is allocated to edge server m and bandwidth B_{jm} is allocated, the data delay can be obtained as follows:

$$t_{jm}^{\text{data}} = \frac{D_{jk}}{B_{jm}}. \tag{1}$$

In the transmission of backjourney data, the return data is usually the processing result, and the backjourney data delay is not considered. Therefore, the backjourney transmission delay only considers the link delay.

The total transmission delay of a task can be expressed as follows:

$$t_{jm}^{\text{tran}} = 2 * t_{jm}^{\text{link}} + \frac{D_{jk}}{B_{jm}}. \tag{2}$$

3.3. Computing Delay Model. θ_{jk} represents the calculation strength required for task k unit data of device j in the current time slot. Therefore, when the task K_j is assigned to edge server m, the total calculation strength by the task processing data is expressed as follows:

$$C = \theta_{jk} * D_{jk}. \tag{3}$$

The computing delay required to process the task can be expressed as follows:

$$t_{jm}^{\text{comput}} = t_k^{\text{load}} + \frac{\theta_{jk} * D_{jk}}{C_{jm}}. \tag{4}$$

3.4. Problem Formulation. In this paper, the binary array x_{jm} represents the task assignment strategy. It can be expressed as follows:

$$\sum_{j \epsilon J} \sum_{m \in M} x_{jm} = \begin{cases} 1, & \text{if } K_j \text{ is assigned to edge server } m, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

In summary, the minimum cost of the system can be expressed as P1 in T_n.

$$\text{P1} \quad \min \quad \sum_{j \epsilon J} \sum_{m \in M} x_{jm} \left(t_{jm}^{\text{tran}} + t_{jm}^{\text{comput}} \right) \tag{6}$$

$$\text{s.t.} \quad \sum_{j \epsilon J} \sum_{m \in M} x_{jm} = 1, \tag{7}$$

$$0 \le \sum_{j \epsilon J} \sum_{m \in M} x_{jm}\left(B_{jm}\right) \le B_m^{\text{idle}}, \tag{8}$$

$$0 \le \sum_{j \epsilon J} \sum_{m \in M} x_{jm}\left(C_{jm}\right) \le C_m^{\text{idle}}, \tag{9}$$

$$2 * t_{jm}^{\text{link}} + \frac{D_{jk}}{B_{jm}} + t_k^{\text{load}} + \frac{\theta_{jk} * D_{jk}}{C_{jm}} \le \tau_{jk}^{\text{worse}}. \tag{10}$$

Here, constraint (7) ensures that the tasks on each device are indivisible and can only be assigned to one edge server. Constraint (8) indicates that the total bandwidth of the tasks assigned on each edge server is not greater than the remaining bandwidth of the current server. Constraint (9) indicates that the total calculation strength of the tasks assigned on each edge server is not greater than the remaining calculationstrength of the current server. Constraint (10) indicates that after all tasks are assigned, the total delay of each task is smaller than the maximum tolerance time.

P1 is a workload allocation scheme that combines task assignment with resource allocation. P1 has proved that it cannot be solved in polynomial time and is proved to be a NP hard problem [15].

4. Proposed Solution

Since P1 is a NP hard problem, in order to simplify the problem and obtain the workload allocation scheme in the fastest way, we propose AP algorithm, which creatively adds historical analysis into the consideration of the problem. In the current time slot, the application to be used in the next time slot is predicted, and the offline algorithm is adopted to preallocate edge resources. Second, the offline problem is decomposed into two subproblems: the task assignment subproblem and the resource allocation subproblem. Piecewise solving can reduce the difficulty of solving and get the suboptimal solution which is close to the optimal solution most quickly.

The edge network workload allocation algorithm based on application prediction proposed in this paper is mainly divided into two stages, as shown in Figure 2.

4.1. Application Prediction. In order to adapt to the prediction scheme in the edge computing scenario, we propose an AP algorithm based on LSTM. In the LSTM prediction method, "input gate, forgetting gate" and control parameter C_t are introduced in each neural unit.

The forgetting gate is shown as follows:

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right). \tag{11}$$

σ is the sigmoid function.

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{12}$$

In the domain of $(-\infty, \infty)$, the value is [-1, 1]. W_f is the weight vector, h_{t-1} is the previous output of the iterative computing process, x_t is the current input sequence matrix, and b_f is the bias vector. f_t is used for the subsequent computing with the control parameter C_{t-1} to determine which kind of information should be discarded.

The input gate is shown as follows:

$$\begin{aligned} I_t &= \sigma(W_I[h_{t-1}, x_t] + b_I), \\ C_t' &= \tan h(W_c[h_{t-1}, x_t] + b_c). \end{aligned} \tag{13}$$

I_t represents new information to be retained, W_I is the weight vector of the input gate, b_I is the bias vector of the input gate. C_t' is the output state of the input gate, W_c is the weight vector of input gate output state, and b_c is the bias vector of input gate output state.

Update the new control vector according to

$$C_t = f_t * C_{t-1} + I_t * C_t'. \tag{14}$$

The result of the output gate is expressed as follows:

$$\begin{aligned} o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o), \\ h_t &= o_t * \tan h(C_t). \end{aligned} \tag{15}$$

o_t is the output gate, W_o is the weight vector of the output gate, and b_o is the bias vector of the output gate. h_t is the output of the output gate, used to calculate the next neuron.

The specific steps of the AP algorithm can be shown in Algorithm 1.

Algorithm 1 is summarized as follows:

(1) The edge network server has the network request usage record of access device. First, the network records are obtained as a reference for historical information

(2) The time information recorded in the network is used to process the data. The network records of each access device are sorted from time information data to time series data. Filter the data into a sequence with the following characteristics; each line includes time, device ID, application ID, and size of the request data volume

(3) The extracted time stamp information (including month, week, hour, minute) is encoded by one-hot as a time feature

(4) The historical data are classified according to the ratio of 8 : 2 to the training set and the test set. Meanwhile, Adam algorithm is used to optimize the parameters and the loss function is calculated by Cross Entropy Loss (CEL)

(5) The AP algorithm based on LSTM is used to train the application prediction model, and the network history information obtained in the previous step is used as the input training set to get the optimal training model

In this paper, we use the data set collected by the application layer of a stable edge system which is published in the

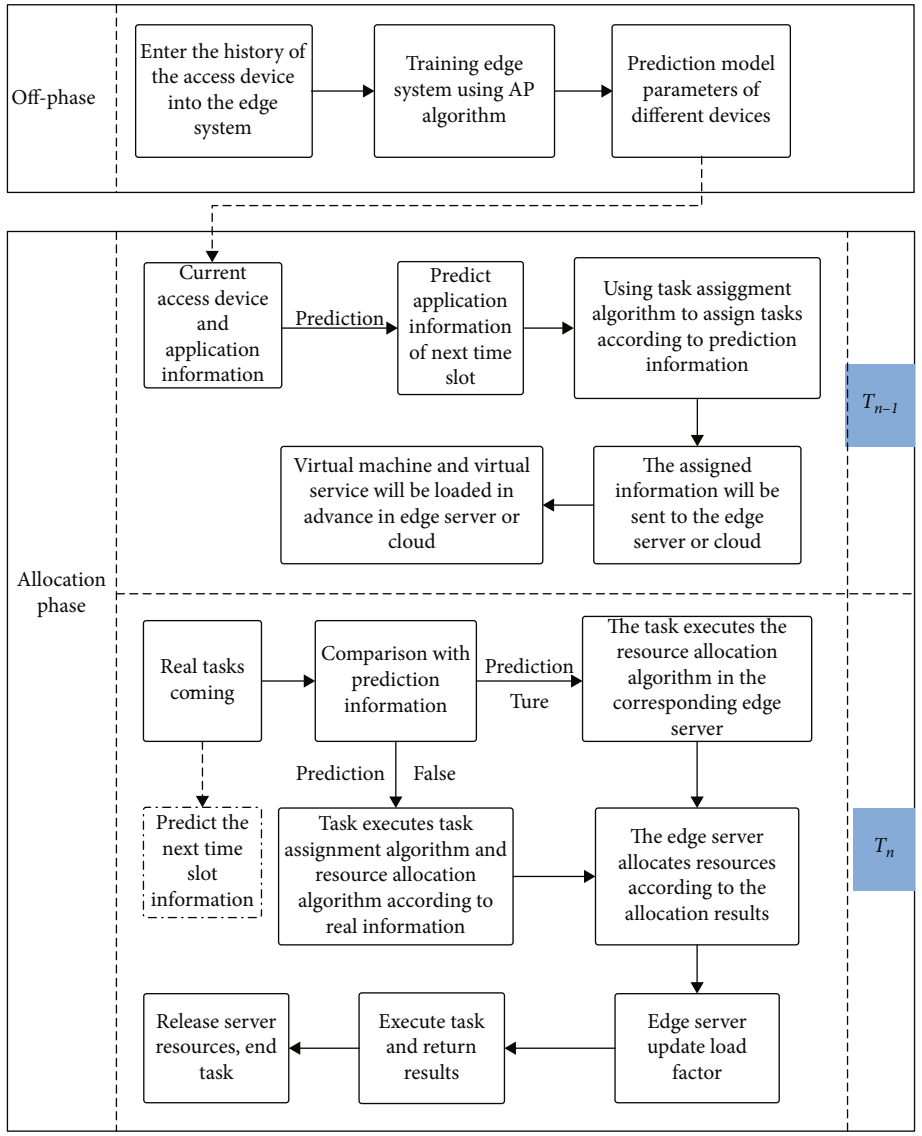


Figure 2: Algorithm execution flow diagram.

Input: the access device network request usage record as H.
Output: Model parameters set P_j of each device $j \in J$

1: Process data H into time series data H'
2: Separate H' by different device j as H'_j
3: **for** each $j \in J$**do**
4: Extract timestamp information from H'_j as feature;
5: Conduct application and different feature into one-hot code;
6: Divide the data H'_j into a training set and a test set according to 8: 2;
7: Train LSTM model;
7: Use Adam algorithm to optimize parameters P_j;
8: **end for**
9: **Return**P_j

Algorithm 1: The AP algorithm based on LSTM.

Kaggle website. The data set contains about 50 access devices and data usage in a base station system. The application usage of different equipment types in the data set is extracted, and different models are trained with the extracted data set. These models are tested, and the average accuracy is obtained. First, in order to compare the performance of different LSTM model complexities, we test the number of layers of different LSTM hidden layers, which are represented by LSTM 1, LSTM 2, LSTM 3, and LSTM 4, representing one layer, two layers, three layers, and four layers of LSTM hidden layers, respectively.

The final comparison results are shown in Table 2. With the increase of the number of LSTM hidden layers, the accuracy of the AP algorithm is also increasing, which is conducive to learning the nonlinear relationship of sample features. However, after increasing from three layers to four layers, the learning ability of model is gradually saturated, and the performance is no longer improved. If we continue to increase the hidden layer, the complexity will increase, which will lead to too many redundant operations. Therefore, under the conditions of this paper, after weighing the complexity and accuracy, the complexity of the LSTM model will adopt three hidden layers. Second, to verify the accuracy of the application prediction model, we compare the LSTM algorithm with the prediction model based on Markov and the prediction model based on RNN. F-Measure parameters are introduced for horizontal comparison. F-Measure parameters are weighted harmonic average of precision and recall, as shown in the formula, which are used to evaluate the merits and demerits of the classification model.

$$F\text{-Measure} = \left(1 + \alpha^2\right) \frac{\text{precision} * \text{recall}}{\alpha^2 * \text{precision} + \text{recall}}, \tag{16}$$

where α is usually equal to 1 and F-Measure is also regarded as F1. In F-Measure of multiclassification problem, the multiclassification problem is usually transformed into N binary classification problems. The F1 of each binary classification can be calculated. The average value of NF1 can obtain the macro F1 of the whole model, which can be used as the evaluation standard of multiple classification models. The higher the value of macro F1 is, the more accurate the model is and the better the performance is. The final comparison results are shown in Table 3, which prove that the prediction model based on LSTM is more accurate.

By using the application prediction model, P1 is divided into two subproblems, namely, task assignment subproblem and resource allocation subproblem.

4.2. Task Assignment Subproblem. In T_n, the arrived tasks form a task set. Each device uses the AP algorithm to predict the application that will be used in T_{n+1}, forming the prediction task set. In T_n, the prediction task set of T_{n+1} is preallocated. In the preallocated resources, the amount of data for each task is estimated based on the historical record, and the computing resources and bandwidth required to meet its delay constraints are given.

Therefore, the value of D_{jk}, B_{jm}, θ_{jk}, and C_{jm} can be determined in formula (6), so it can be converted into a constraint

Table 2: Comparison of accuracy of different model layers.

Model layers	Accuracy
LSTM 1	78.61%
LSTM 2	84.20%
LSTM 3	87.20%
LSTM 4	87.20%

Table 3: Comparison of performance of different models.

Model	Accuracy rate	Macro F1
Markov	65.53%	0.52
LF-Markov	76.35%	0.61
RNN	82.30%	0.67
LSTM	87.20%	0.74

condition. Owing to the preallocation, the service load time t_k^{load} is already loaded before the next time slot arrives and is not counted in the calculation. Then, the variables are only the link delay t_{jm}^{link}, which is between the device j and server m , and the task assignment strategy x_{jm}.

Given the prediction task set, the task assignment subproblem in T_{n+1} is determined by the link delay and task assignment strategy and thus can be obtained by solving the following problem:

$$\begin{aligned}
\text{P2} \quad \min \quad & \sum_{j\epsilon J} \sum_{m\in M} x_{jm}\left(2 * t_{jm}^{\text{link}}\right) \\
\text{s.t.} \quad & \sum_{j\epsilon J} \sum_{m\in M} x_{jm} = 1, \\
& 0 \le \sum_{j\epsilon J} \sum_{m\in M} x_{jm}\left(B_{jm}\right) \le B_m^{\text{idle}}, \\
& 0 \le \sum_{j\epsilon J} \sum_{m\in M} x_{jm}\left(C_{jm}\right) \le C_m^{\text{idle}}, \\
& 2 * t_{jm}^{\text{link}} \le \tau_{jk}^{\text{worse}} - \frac{D_{jk}}{B_{jm}} - \frac{\theta_{jk} * D_{jk}}{C_{jm}}.
\end{aligned} \tag{17}$$

In P2, all variables represent the predictive variables of the next time slot. P2 is converted to a packing problem. A descending best fit algorithm based on greedy thought is adopted to solve this problem.

Algorithm 2 shows that the tasks to be assigned are arranged in a descending order according to the resource consumption. Since there are two constraint variables, the tasks are arranged according to the amount of data. In order to find the edge server that can meet its needs, the one with the lowest link delay is selected to assign the tasks to the edge server.

Algorithm 2 is summarized as the following steps:

(1) The set Z of prediction application K_j for the next time slot and the corresponding prediction information tuple $(D_{jk}, B_{jm}, \theta_{jk}, C_{jm})$ is input

Input: Z, M, and the estimated information tuple $(D_{jk}, B_{jm}, \theta_{jk}, C_{jm})$ of K_j
Output: Task Allocation plan x_{jm}
1: **Initialization:** $\sum_{j\epsilon J}\sum_{m\epsilon M} x_{jm} = 0$
2: Given the predicted application set of next time slot Z, sort K_j by D_{jk} from largest to smallest as Z'
3: **for** $K_j \in Z'$ **do**
4: **for** $m \in M$ **do**
5: If B_{jm} and C_{jm} can be satisfied by m, add to temp set m'
6: **end for**
7: **for** $m \in m'$ **do**
8: Find the lowest t_{jm}^{link}, which represents transmission link delay between device j and server m, set $x_{jm} = 1$
9: **end for**
10: Clear m' and remove K_j from Z'
11: Update C_m^{idle}, B_m^{idle}, μ_m^B and μ_m^C of server m
12: **end for**
13: If $Z' \neq \varnothing$ send Z' and Information tuple to cloud center
14: **Return** x_{jm}

Algorithm 2: Task assignment.

(2) The tasks to be assigned are in a descending order according to the amount of data D_{jk}

(3) By finding all the edge servers whose remaining bandwidth and computing resources satisfy B_{jm} and C_{jm}, the tasks in a descending order are unloaded to the server with the lowest link delay, and then, the remaining bandwidth and computing resources of the server, as well as the load factors μ_m^B and μ_m^C, are updated. After all tasks are assigned, if the task set is not empty, it proves that the current task set has exceeded the edge server's load capacity, and the remaining tasks are assigned to the cloud for execution

After Algorithm 2 is executed, the task assignment scheme of all tasks will be output.

4.3. Resource Allocation Subproblem. After determining the task assignment scheme, x_{jm} is determined. The constraint formula can be expressed as P3 in each edge server m.

$$\text{P3} \quad \min \quad \sum_{j\epsilon J}\sum_{m\epsilon M} x_{jm}\left(2 * t_{jm}^{\text{link}} + \frac{D_{jk}}{B_{jm}} + \frac{\theta_{jk} * D_{jk}}{C_{jm}}\right) \tag{18}$$

$$\text{s.t.} \quad 0 \le \sum_{j\epsilon J}\sum_{m\epsilon M} x_{jm}\left(B_{jm}\right) \le B_m^{\text{idle}}, \tag{19}$$

$$0 \le \sum_{j\epsilon J}\sum_{m\epsilon M} x_{jm}\left(C_{jm}\right) \le C_m^{\text{idle}}, \tag{20}$$

$$0 \le 2 * t_{jm}^{\text{link}} + \frac{D_{jk}}{B_{jm}} + \frac{\theta_{jk} * D_{jk}}{C_{jm}} \le \tau_{jk}^{\text{worse}}. \tag{21}$$

In solving this problem, there are two optimization variables, B_{jm} and C_{jm}. Based on the preallocated computing and bandwidth resources, the ratio of bandwidth resources and computing resources of edge server m is calculated μ_m^B and μ_m^C, also known as load factor. In general, the load time of virtual service increases with the increase of load factor.

$$\begin{aligned} \mu_m^B &= \sum_{j\in J} \frac{x_{jm}\left(B_{jm}\right)}{B_m}, \\ \mu_m^C &= \sum_{j\in J} \frac{x_{jm}\left(C_{jm}\right)}{C_m}. \end{aligned} \tag{22}$$

Compared with μ_m^B and μ_m^C, the parameter with high occupancy rate is determined as the main constraint of delay. For example, when $\mu_m^B = 0.2$ and $\mu_m^C = 0.6$, it can be determined that the current computing resource is the main constraint of delay. Then, the parameter of preallocated computing resource is taken into P3, and C_{jm} is used as the known condition to optimize the allocation of B_{jm}.

In general, the update process of the algorithm needs to avoid server overload and the prediction error leading to resource reallocation. Therefore, according to prior experience, $\mu_m^C \le 0.7$ and $\mu_m^B \le 0.8$.

When a parameter is fixed, the second partial derivative of formula (18) is greater than 0, the Hessian matrix is a positive definite matrix. Moreover, the constraints of P3 is transformed into linear constraints; P3 is a convex programming problem.

As P3 is a convex problem, we can derive the optimal solution of P3 by solving the KKT condition [21]. Algorithm 3 can be expressed as the following steps.

(1) By traversing each server, μ_m^B and μ_m^C of the current server are calculated according to the predicted bandwidth and computing resources

Input: x_{jm}, the predicted bandwidth B'_{jm}, and predicted computing resources C'_{jm}
Output: B_{jm}, C_{jm}
1: **for**m $\in M$**do**
2: Add all B'_{jm} to calculate μ^B_m;
3: Add all C'_{jm} to calculate μ^C_m;
4: **if**$\mu^B_m \geq \mu^C_m$**then**
5: Let $B_{jm} \leftarrow B'_{jm}$;
6: Set formula (18) Lagrange function on C_{jm}, get $L(C_{jm}, \lambda, \mu)$;
7: Calculate KKT condition, get the result set C_{jm};
8: **else**
9: Let $C_{jm} \leftarrow C'_{jm}$;
10: Set formula (18) Lagrange function on B_{jm}, get $L(B_{jm}, \lambda, \mu)$;
11: Calculate KKT condition, get the result set B_{jm};
11: **end if**
12: **end for**
13: **Return**B_{jm}, C_{jm}.

Algorithm 3: Resource allocation.

(2) The constraints of the current server task are determined by comparing μ^B_m and μ^C_m

(3) According to the comparison results, the fixed parameters are determined and the corresponding Lagrange functions are obtained

(4) It calculates the KKT condition and gets the result

5. Simulation

5.1. Simulation Setting. In this section, we set up simulations of the proposed scheme to evaluate its performance, which uses iFogSim simulation platform [22]. In parameter setting, the simulation environment consists of 5 edge servers, where the assignable bandwidth of each edge server is 1000 Mbps and the CPU frequency of each edge server is 1 Ghz. Meanwhile, the length of each time slot is set as 1 s. According to prior experience, $\mu^C_m \leq 0.7$ and $\mu^B_m \leq 0.8$. Moreover, the link delay between different devices and edge servers is randomly chosen according to the normal distribution. The link delay connected to different devices is different. The size of the link delay is set within 20 ms. In addition, the loading time of virtual service among edge server is only related to the application types, and the loading time of virtual service for different applications is set between 20 ms and 50 ms. The number of connected devices of the edge system is set between 0 and 50.

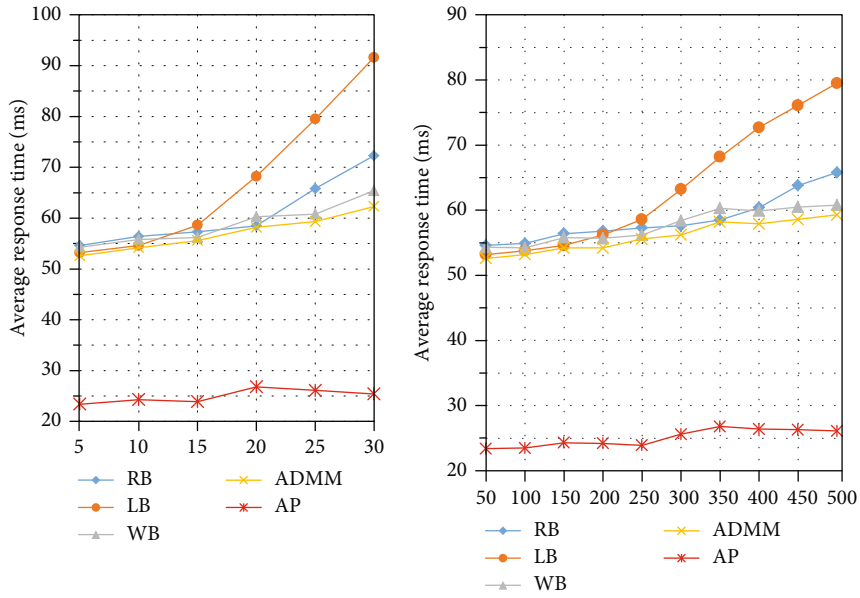
We select four other workload allocation strategies for comparison: (1) the random-based strategy (RB), (2) the latency-based strategy (LB), (3) the weighted-based strategy (WB) [11], and (4) the ADMM algorithm based on cooperative task offloading [23]. The basic idea of RB is that tasks will be randomly assigned to different computing servers after they arrive, and a competition scheme of preemptive resources is adopted. LB is used to select the link with the lowest delay according to different link delays when a task arrives and then offload the task to the edge server. WB is used to set different weights for different tasks. The weight is set according to the delay sensitivity of tasks. Through the optimization iteration of weights, the workload allocation scheme is optimized. The ADMM algorithm based on cooperative task offloading is a series of reconstructions through the reconstruction linearization technique, and a parallel optimization framework is proposed by using ADMM and difference of convex function (DC) programming to solve the problem.

5.2. Comparison of Average Response Time. We discuss the impact of the number of access devices and network traffic on the average response time of tasks (also known as the average wait time of tasks). Task response time is defined as the time between the submission of a task request to the edge server, the edge server setting up the virtual machine, and the notification of the device to start offloading task data. Fast response times are critical to improving user QoE.

Figure 3(a) shows the average response time for different numbers of access devices, in which AP achieves lower response time as compared to the other four algorithms. In Figure 3(b), the average response time for different network traffic in AP is significantly smaller than that of RB, LB, WB, and ADMM.

As shown in Figure 3, when the number of tasks increases, the average response time of LB grows the fastest. When the link delay of a server to most devices is the lowest, it is easy for LB to cause users to gather in a specific server. Therefore, when the access devices are increased, LB has a large probability to overload a single server, which affects the virtual machine establishment process, resulting in a large increase in the average response time. The average response time increases from 53.2 ms for the first 5 access devices to 91.6 ms for the 30 access devices.

RB can largely avoid server overload. However, the task on each device cannot be assigned to the optimal server node, which leads to the increase of link delay. In both experiments,



(a) Total number of access equipment

(b) Total network traffic per slot (M)

Figure 3: Comparison of average response time of different algorithms.

it can be seen that when the number of devices increases, the growth trend is slightly smaller than that of LB.

WB is a workload allocation strategy based on different weights, so its performance is better than LB and RB. When the number of access devices initially increases, the response time does not change significantly. When the access devices reach 30, the average response time only increases from 54.3 ms to 65.4 ms.

The task assignment of the ADMM algorithm is close to the theoretical optimal performance, so it can allocate tasks optimally. When the number of access devices increases, the response time change is not obvious, which can have better performance than WB. When the access devices reach 30, the average response time increases from 52.6 ms to 62.3 ms.

AP algorithm is proposed in this paper, and it can be seen that the average response time is greatly reduced. Since the average response time mainly depends on link delay and server load time, the AP algorithm in this paper has loaded the virtual machine of the application on the server in advance, so the server load time can be saved. Theoretically, the average response time of the AP algorithm is only affected by link delay and prediction accuracy. It can be seen that when the access devices reach 30, the response time is only 25.4 ms, which can still reduce the response delay by about 60% compared with the current best performance ADMM algorithm.

5.3. Comparison of Average Completion Time. The average completion time is defined as the time interval from the task request submitted to the edge server until the server returns the final result. It is a direct index to measure the performance of workload allocation scheme.

5.3.1. The Impact of Different Access Devices. In parameter setting, in order to better compare the performance of different strategies, we set the task data size within 20 M per time slot without extreme data. The calculation strength θ_{jk} required for each task unit data is generated between $1 * 10^7$ CPU cycle/M and $2 * 10^7$ CPU cycles/M. The impact of energy consumption on transmission delay is ignored.

Comparing average completion time with different numbers of access devices under different algorithms in Figure 4, it can be seen that the completion time increases with the increase of the number of access devices. When 5 devices are connected, each server node is in no load state, and the computing and bandwidth resources are sufficient, so the transmission delay is the main part of the total completion time of the task. The performance of LB, WB, and ADMM algorithms is relatively consistent, and the average completion time is between 560 ms and 580 ms. The average completion time of the RB algorithm is higher than that of other algorithms due to the high link delay of some tasks, which is 594.6 ms.

With the increase of access devices, the LB algorithm may cause some server overload, resulting in the average completion time higher than other algorithms. Therefore, the RB algorithm performs better than the LB algorithm in small and medium task load.

The overall performance of the WB algorithm is better than the RB algorithm and LB algorithm. When the number of tasks increases from small to large scale, the average completion time increases within a reasonable range, close to linear growth. Since the RB algorithm and LB algorithm are close to exponential growth, user QoE drops sharply. The RB algorithm has the worst performance. When the access device reaches 30, the average task completion time of the RB algorithm has increased to 952.3 ms.

The ADMM algorithm is close to the theoretical optimal value. When the number of tasks is small, the average task completion time almost does not increase. When the number

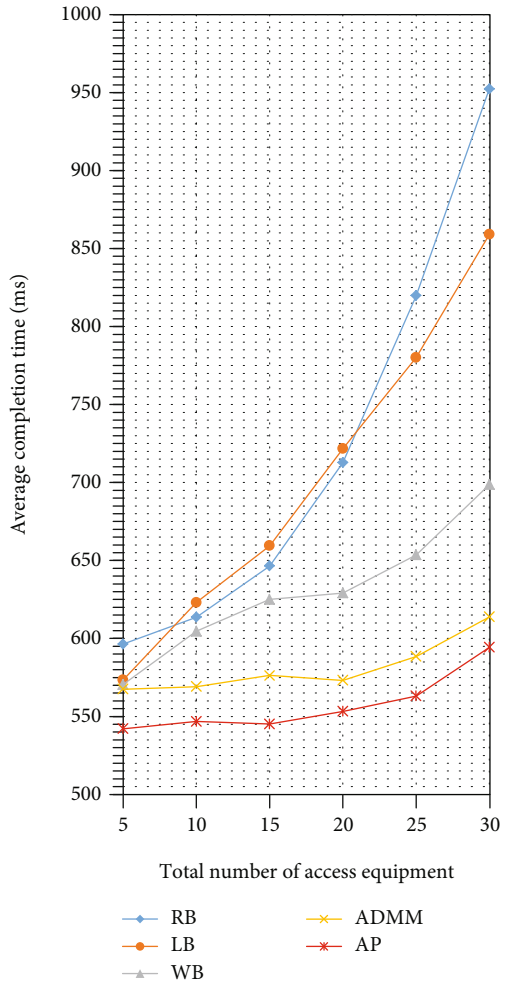


Figure 4: Comparison of average completion time of different access devices.

of tasks increases to a large scale, it only shows a small increase. Meanwhile, owing to ignoring service load delay by application prediction, the AP algorithm is better than the ADMM algorithm.

5.3.2. The Impact of Different Server Processing Capabilities. In order to further compare and explore the performance of different algorithms, we fix 15 access devices in the experiment, which are in the state of medium load. By changing the processing capacity of each server, the average completion time curve is obtained as shown in Figure 5. It can be seen that the average completion time curve decreases significantly between 0.6 and 1.0 Ghz in processing capacity. After 1.0 Ghz, with the increase of processing capacity, the average completion time decreases less. Thus, the server can still be under normal load when the processing capacity is around 1.0 Ghz. When the processing capacity is reduced, each server is gradually full or even overloaded, and the average completion time increases sharply. When the load capacity is gradually increased, the impact on the average completion time is not significant.

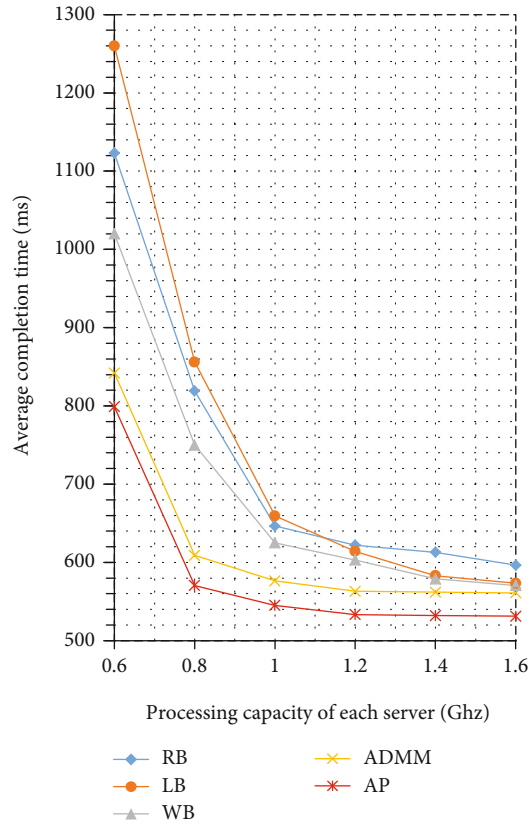


Figure 5: Comparison of average completion time of servers under different processing capabilities.

In the comparison of the average completion curves of the five algorithms, it can be seen that the LB algorithm has the worst performance when the server processing capacity is insufficient. The RB algorithm has the worst performance when the server performance is sufficient. The WB algorithm is slightly better than the LB algorithm and RB algorithm. However, the AP algorithm proposed in this paper has the optimal performance under different server processing capacities. The average completion time is lower than that of the ADMM algorithm between 20 ms and 50 ms, which is close to the theoretical performance.

5.3.3. The Impact of Different Network Traffic. Figure 6 shows the average completion time curve of different algorithms with the change of network traffic in a day.

In the experiment, we select one day's data from the data set and input them into the simulation platform. The abscissa unit is hour, 1 represents the data between 0 and 1. A total of 24 hours of data is counted. The network traffic represents the peak of network traffic within an hour. The ordinate represents the average completion time when network peak occurs.

It can be seen that the RB algorithm and LB algorithm are greatly affected by the network traffic. When the network traffic increases, the average completion time will also increase significantly. The performance of the WB algorithm is better than the RB and LB algorithm, but worse than the

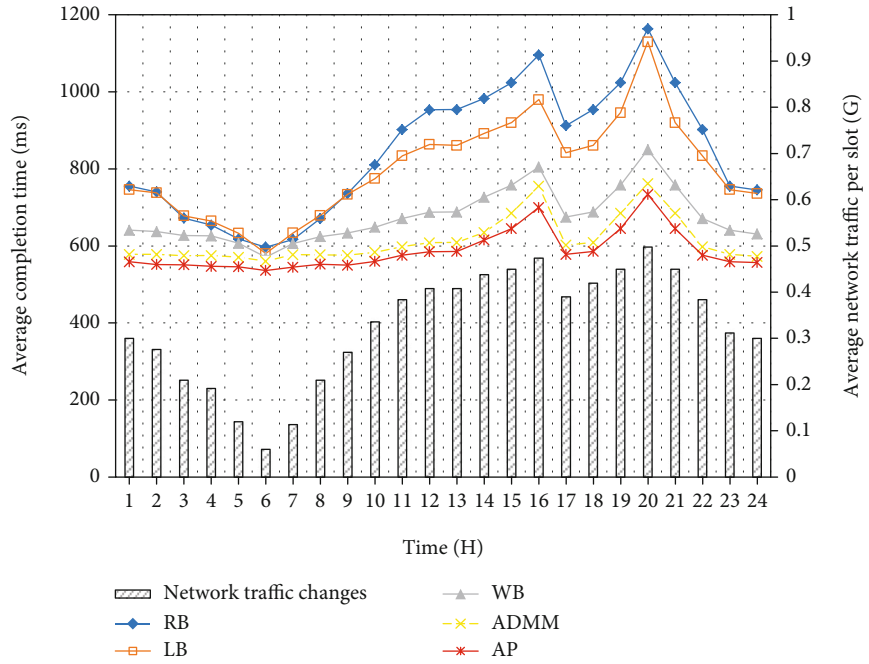


Figure 6: Comparison of average completion time under network traffic changes.

ADMM algorithm. When the network traffic is below 0.55 G, with the change of network traffic, the average task completion time does not change much, ensuring a certain stability. The AP algorithm proposed in this paper, due to the preloading of the virtual machine, reduces the time of virtual machine establishment process for most tasks and shows the optimal performance.

The experiment results confirm the feasibility of the AP algorithm. Meanwhile, when the traffic increases to a large load, the growth of the average completion time is the same as that of the ADMM algorithm, which proves that the AP algorithm in this paper is also close to the theoretical optimal result in workload allocation.

6. Conclusions

In this paper, we propose the edge workload allocation scheme using the AP algorithm. By using the AP algorithm, the workload allocation problem is divided into two subproblems, which are the task assignment subproblem and the resource allocation subproblem. The task assignment subproblem can be transformed into a packing problem and solved by the descending best fit algorithm. Based on the task assignment scheme, the resource allocation subproblem can be transformed into a convex programming problem by changing the constraint conditions and solved by KKT condition. The simulation results show that compared with the current best performance ADMM algorithm, the AP algorithm proposed in this paper can reduce the response delay by about 60%. In terms of average completion time, compared with the current best performance ADMM algorithm, the performance is still improved by 5%-9%, which has high practical significance.

Data Availability

The simulation data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The work was supported by "National Natural Science Foundation of China" with No. 61902052, "National Key Research and Development Plan" with No. 2017YFC0821003-2, "Dalian Science and Technology Innovation Fund" with Nos. 2019J11CY004 and 2020JJ26GX037, and "the Fundamental Research Funds for the Central Universities" with Nos. DUT19RC(3)003 and DUT20ZD210.

References

[1] D. Reinsel, J. Gantz, and J. Rydning, "Data age 2025: the digitization of the world from edge to core," pp. 1–29, 2018, IDC White Paper Doc#US44413318.

[2] W. Li, Z. Chen, X. Gao, W. Liu, and J. Wang, "Multimodel framework for indoor localization under mobile edge computing environment," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4844–4853, 2019.

[3] A. Yousefpour, C. Fung, T. Nguyen et al., "All one needs to know about fog computing and related edge computing paradigms: a complete survey," *Journal of Systems Architecture*, vol. 98, pp. 289–330, 2019.

[4] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: a survey," *Future Generation Computer Systems*, vol. 97, pp. 219–235, 2019.

[5] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[6] Y. Mao, J. Zhang, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3590–3605, 2016.

[7] X. Lyu, W. Ni, H. Tian et al., "Optimal schedule of mobile edge computing for Internet of Things using partial information," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2606–2615, 2017.

[8] M.-H. Chen, M. Dong, and B. Liang, "Resource sharing of a computing access point for multi-user mobile cloud offloading with delay constraints," *IEEE Transactions on Mobile Computing*, vol. 17, no. 12, pp. 2868–2881, 2018.

[9] J. du, L. Zhao, J. Feng, and X. Chu, "Computation offloading and resource allocation in mixed fog/cloud computing systems with min-max fairness guarantee," *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1594–1608, 2018.

[10] L. Liu, Z. Chang, X. Guo, S. Mao, and T. Ristaniemi, "Multiobjective optimization for computation offloading in fog computing," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 283–294, 2018.

[11] H. Tan, Z. Han, X.-Y. Li, and F. C. M. Lau, "Online job dispatching and scheduling in edge-clouds," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, Atlanta, GA, USA, 2017.

[12] Y. Sun, S. Zhou, and J. Xu, "EMM: energy-aware mobility management for mobile edge computing in ultra dense networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 11, pp. 2637–2646, 2017.

[13] M. Jia, J. Cao, and W. Liang, "Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks," *IEEE Transactions on Cloud Computing*, vol. 5, no. 4, pp. 725–737, 2017.

[14] Y. Xiao and M. Krunz, "QoE and power efficiency tradeoff for fog computing networks with fog node cooperation," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, Atlanta, GA, USA, 2017.

[15] Q. Fan and N. Ansari, "Application aware workload allocation for edge computing-based IoT," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2146–2153, 2018.

[16] T. Dlamini and A. F. Gambin, "Adaptive resource management for a virtualized computing platform within edge computing," in *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, Boston, MA, USA, 2019.

[17] A. Acharya, *Edge-Assisted Workload-Aware Image Processing System*, Boise State University Theses and Dissertations, 2019.

[18] J. Xu, L. Chen, and Z. Pan, "Joint service caching and task offloading for mobile edge computing in dense networks," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, Honolulu, HI, USA, 2018.

[19] C. Shu, Z. Zhao, Y. Han, and G. Min, "Dependency-aware and latency-optimal computation offloading for multi-user edge computing networks," in *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, Boston, MA, USA, 2019.

[20] P. Mach and Z. Becvar, "Mobile edge computing: a survey on architecture and computation offloading," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017, thirdquarter.

[21] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge university press, 2013.

[22] H. Gupta, A. Vahid Dastjerdi, S. K. Ghosh, and R. Buyya, "iFogSim: a toolkit for modeling and simulation of resource management techniques in the Internet of Things, edge and fog computing environments," *Software: Practice and Experience*, vol. 47, no. 9, pp. 1275–1296, 2017.

[23] Y. Wang, X. Tao, X. Zhang, P. Zhang, and Y. T. Hou, "Cooperative task offloading in three-tier mobile computing networks: an ADMM framework," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2763–2776, 2019.