

Retraction

Retracted: Application of Data Mining in Network Information Dynamic Push Software

Security and Communication Networks

Received 25 July 2023; Accepted 25 July 2023; Published 26 July 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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

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

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

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

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

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

References

- [1] X. Hong, "Application of Data Mining in Network Information Dynamic Push Software," *Security and Communication Networks*, vol. 2022, Article ID 3689802, 7 pages, 2022.

Research Article

Application of Data Mining in Network Information Dynamic Push Software

Xiaoyi Hong 

Xinxiang Vocational and Technical College, Xinxiang, Henan 453003, China

Correspondence should be addressed to Xiaoyi Hong; 201903328@stu.ncwu.edu.cn

Received 26 June 2022; Revised 21 July 2022; Accepted 27 July 2022; Published 10 August 2022

Academic Editor: C. Venkatesan

Copyright © 2022 Xiaoyi Hong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to solve the problem that it is difficult for each user to obtain the most needed information due to different backgrounds, a method of data mining network information is proposed. The main steps of this method are as follows: (1) designing data mining rules based on association rules; (2) screening candidate sets of network information data mining; (3) candidate set information data mining. The experimental results show that the p value of this method is more than 90.0% and that of the traditional method is only within the range of 30.0%~70.0%. Obviously, the p value of this method is 20.0%~60.0% higher than that of the traditional method. This method can effectively solve the problem that it is difficult for each user to obtain the most needed information due to background differences.

1. Introduction

With the advent of the information age, the Internet has become a global super database. The rich resource information has broadened people's vision. However, due to the scattered distribution, dynamic changes, and complex structure of Internet information, information overload and information confusion have become social problems that hinder the efficiency of the Internet. Questions such as how to efficiently and comprehensively obtain the required information from the massive information and how to improve the active information service ability of the network and meet the personalized needs of users have always been the hot topics of information consulting experts. With the rapid development of the Internet, the Internet has become the fastest, most convenient, and most effective medium for people to transmit information. Compared with traditional media, the advantage of the Internet lies in its strong interactivity and vividness. When users browse web pages or browse social software trends, various forms of push will appear, providing users with a variety of information and enriching users' lives to a certain extent. Due to the different backgrounds of each user, the accompanying problem is that users are increasingly difficult to obtain the information they

need most [1, 2]. In this environment, with the rise and popularity of intelligent mobile terminals, it brings great convenience for dynamic push. The information push service can actively enrich users' lives to a certain extent according to their needs, transmit the latest information to the corresponding user devices by categories, and solve the problems that it is increasingly difficult for users to obtain the information they need most [3].

2. Literature Review

With the rapid development of Internet, the network information resources have increased rapidly, and the problem of network information overload has become increasingly prominent. The network information retrieval system represented by Yahoo has emerged and developed rapidly. The network information retrieval system is generally composed of a robot, an index database, and a query engine. The information collector robot traverses the WWW to find as much new information as possible; using full-text retrieval technology to index the collected information and store it in the index database can greatly improve the speed of information retrieval; the query engine receives and analyzes the user's query, traverses the index database according to a

relatively simple matching strategy (simple Boolean model or fuzzy Boolean model), and finally submits the result address set to the user [4]. Due to the limitation of the research level of artificial intelligence, robots cannot realize the accurate classification of information at present. Most search sites process the information manually, and the speed of information sorting lags far behind the growth rate of network information. Therefore, the current Chinese and English search engines have low precision and recall, which cannot meet the needs of users for high-quality network information services. Teachers can obtain the characteristics of students' learning behaviors through the teaching management system of the network teaching platform, and can select some important characteristics as the statistical attributes of the network learning behaviors, such as the number of times students log in the course to learn, the completion of the time test of continuous learning, the topics published in the forum area and the information of participation and reply. Quantifying the network learning behavior reasonably, as the data of data statistics and mining, is helpful to establish the learner model and help teachers master the learning characteristics of students.

The attributes of some common e-learning behaviors are shown in Table 1.

Based on the current research, this paper proposes a network information method of data mining. Therefore, it is the application of data mining technology, artificial intelligence information retrieval, and natural language understanding technology in network information processing. Different from traditional information retrieval, data mining can extract the deep-seated information required by users from the extended comparison of concepts and related factors in the database composed of heterogeneous data. It will reform the traditional information service mode and form a new information service combination suitable for the requirements of the network era [5]. The main contents of this method are as follows: (1) design of data mining rules based on association rules; (2) screening candidate sets of network information data mining; (3) candidate set information data mining. The performance of data mining is tested through comparative experiments. This method solves the problem that it is more and more difficult to obtain the most needed information due to the background difference of each user.

3. Research Methods

3.1. Overview of Data Mining. Data mining, also known as knowledge discovery in databases, rose in 1989. It is the product of the integration of knowledge from multiple disciplines, including research achievements in many disciplines such as machine learning, database application technology, statistics, artificial intelligence, and so on [6]. Therefore, data mining is defined as a means of sorting, inducing, and discovering high-value models or data from massive data by using the knowledge and technology of machine learning, statistical learning, and other related aspects, and extracting a novel, effective, potentially useful, and understandable model processing process, as shown in Figure 1.

3.2. Excavation Process

3.2.1. Define the Mining Theme. Before data mining, first define your own data orientation, determine the direction and scope of data mining, and then implement data mining to avoid data redundancy, data deviation, and other problems and avoid blind mining [7].

3.2.2. Data Processing. Data processing is an important link in the process of data mining. Only by ensuring the accuracy and effectiveness of data can we ensure the significance of data mining. This link is divided into three small links, namely, data selection, data preprocessing, and data conversion [8]. (1) Preliminary data preparation. Collect relevant data according to the research topic, classify the collected data, eliminate the data irrelevant to the topic or with large deviation, and leave the data consistent with the topic. (2) Data processing. Secondly, process the sorted data, delete blank fields and meaningless data, and ensure the validity of the remaining data. (3) Data conversion. It is the precondition of data mining to cluster the retained valid data according to the research subject to meet the data mining format requirements.

3.2.3. Data Mining. Data mining is to carry out substantive data mining, then select an algorithm suitable for the data research according to the topic, and then carry out data mining. This link is the core link of data mining, as shown in Figure 2.

3.2.4. Data Analysis. After the completion of data mining, the last step is to explain the research according to the data results mined [9]. Its main role is to determine whether the knowledge model is effective so as to find more meaningful knowledge models, as shown in Figure 3.

3.3. Main Methods

3.3.1. Decision Tree. Decision tree is the mainstream method of data mining. It clearly describes the process of data decision and data classification in the form of tree. This algorithm is relatively simple and intuitive. The decision trees generated in different scenarios are different, so the decision trees are also called classification trees and regression trees. The classical algorithms of decision tree data mining methods are mainly the ID3 algorithm and C4.5 algorithm [10, 11]. The decision tree generation process is shown in Figure 4.

- (1) Cart algorithm [12]. It is a simple binary tree algorithm, which is often used in simple data to generate a simple binary tree.
- (2) ID3 algorithm. ID3 algorithm is a relatively early algorithm in the decision tree algorithm. It finds out the attributes represented by each node in the tree through a series of rules based on the data

TABLE 1: Common e-learning behaviors and attributes.

E-learning behavior	Behavior attribute
Browse the web	Title, website address, subject information, and time (entering and leaving)
Browse lesson plans and knowledge points	Subject information, login frequency, and time (entry and exit)
Watch micro classes and videos	Subject information, name, address, and time (start, end)
Information retrieval	Search keywords, search engines, and search results
Download save	Search keywords, download source, format, content introduction, and file size
Favorite websites	Website address, name, and website introduction
Ask questions	Subject, reading times, and reply times
Question answering	Subject, frequency, correct times, and wrong times
Send and receive mail	Sending and receiving address and e-mail subject
Online discussion	Discussion topic, number of posts read, number of posts posted, frequency, and time
Real time communication	Tools used, communication time, communication object, and communication content
Online test	Online (unit, comprehensive) test time, times, and results
Submit job	Number, time, subject, and grade of submitted assignments
Publication of works	Name of work, author of work, tools used, and publishing website

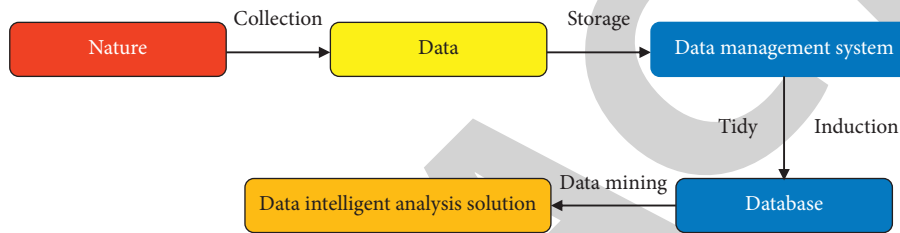


FIGURE 1: Meaning of data mining.

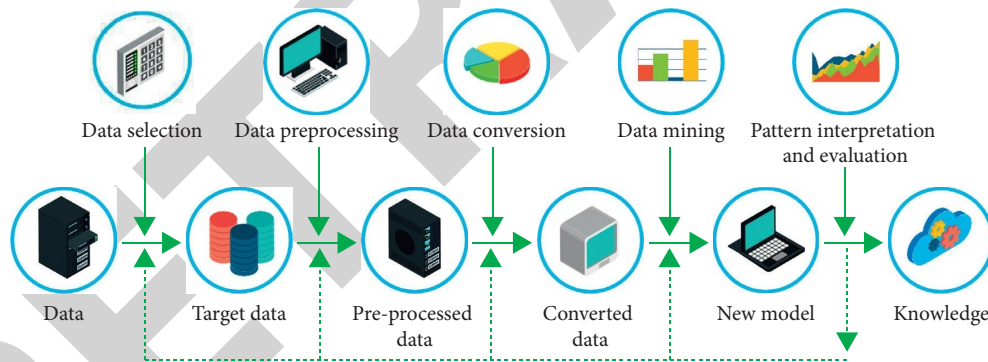


FIGURE 2: Data mining implementation process.

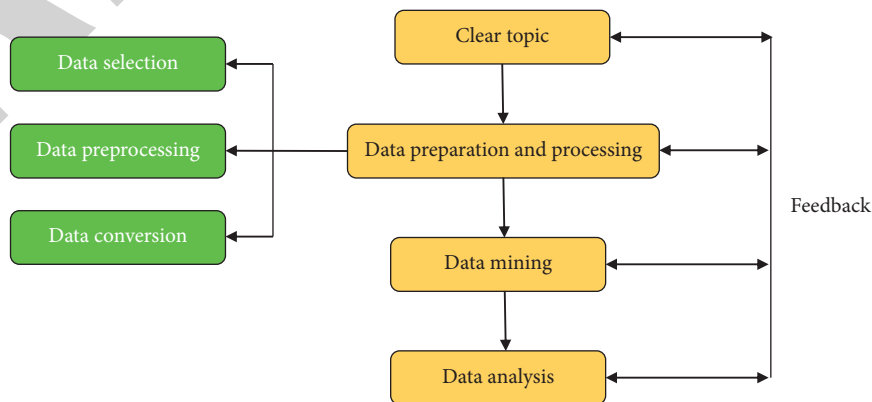


FIGURE 3: Data mining process.

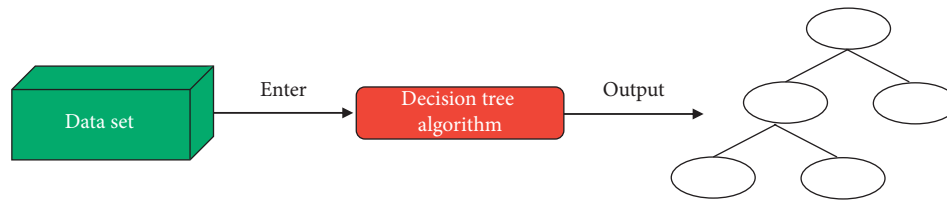


FIGURE 4: Decision tree generation process.

information and finally generates the data into the form of decision tree based on the entropy in the algorithm.

- (3) C4.5 algorithm. C4.5 algorithm is optimized and improved on the basis of ID3 algorithm. This algorithm uses information gain or entropy to optimize the node division process of the decision tree, improve the decision tree, and make the decision tree more friendly [13].

3.3.2. Cluster Analysis. The essence of cluster analysis is to find out the classification basis of data according to the research topic, classify and process the data according to this basis, refine the data into different types of data sets, and ensure that the data in each set has similarity, and there are differences between different sets and then use the data visualization technology to show it to users friendly, which is called cluster analysis [14]. The main algorithm is the k-means algorithm. The outstanding advantage of this algorithm lies in its simple principle and efficient application. It is very suitable for processing large-scale data and has achieved good application results in many fields, including data analysis, personalized recommendation, data classification, image recognition, and so on.

3.3.3. Association Rule Analysis. Association rule analysis is one of the commonly used methods in data mining. There is a mutual relationship between things, and this relationship is called association. Association rules refer to the hidden relationship rules between things, while association rule analysis refers to the process of finding and analyzing the information between the set value association rules between things. Its main algorithm is the Apriori algorithm, which is the most influential algorithm for mining frequent item sets of single-dimensional, single-layer, and Boolean association rules [15]. Although the Apriori algorithm can solve the analysis of corresponding data association rules, it still has some defects. Then, the FP growth algorithm is proposed to make up for the defect of the Apriori algorithm in generating candidate item sets.

3.3.4. Support Vector Machine. The support vector machine (SVM) is a binary classification model. The basic model is a linear classifier with the largest interval defined in the feature space. When the interval is the largest, it is different from the perceptron [16]. The core idea is that support vector samples will play a key role in recognition. The support vector is the sample point closest to the classification hyperplane, which

is the support vector classifier. Through this classification hyperplane, the sample data can be divided into two, as shown in Figure 5.

3.4. Design of Network Information Data Mining Method

3.4.1. Data Mining Rule Design Based on Association Rules.

When mining information data in the network environment, the main purpose of setting data mining rules is to find out the frequent transactions in the massive data sets, that is, frequent item sets. Association rules is a kind of mining algorithm which takes the growth trend as the main form. This paper designs the rules of network information data mining based on association rules. In network information data mining, we need to go through two network databases. For the first time, the candidate sets are mined at the beginning of mining. In this stage, the single frequent item sets generated are the mining results. The second time, in the process of mining candidate sets, the mining data with high complexity are optimized to relieve the pressure in the process of mining execution. The specific mining rules are as follows: first, the selected samples to be mined are processed in blocks, and the processed results are input into each node of the cluster. The support of each data node is calculated through association rules. Then, complete the execution of the map program, obtain the local relevant data sets from the network files, input a known data record in the mapper, use the combiner to complete the simple consolidation of the local data set records, and uniformly allocate the keys and values with the same protective equipment to a reducer [17]. Then, all the extracted data values are accumulated and integrated into a whole, and the support calculated above is combined into a sequence diagram from small to large.

Secondly, input another numerical record information in the mapper, compare it with the numerical record information in the previous step, send the same data information to the same node, mine it frequently, and finally get the corresponding mining results. Finally, the data information of different data values is unified into different data nodes to ensure that the corresponding frequent item sets do not exist in one data node at the same time so as to ensure that the data information after mining has a certain regular order. Combined with the default sorting function of key values in association rules, the key values are replaced by one of the construction algorithms, and all the results are summarized. The data obtained are the final result obtained through data mining.

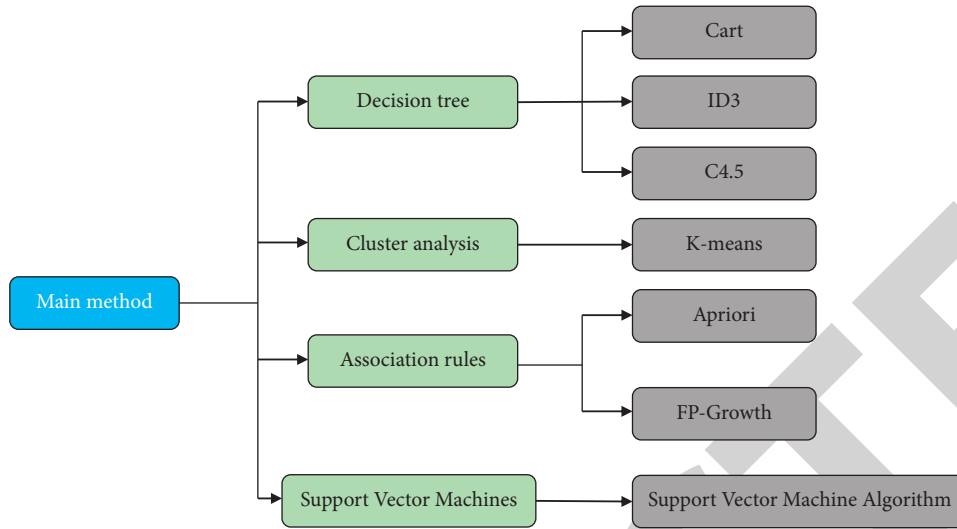


FIGURE 5: Support vector machine classification model.

3.4.2. Filtering Candidate Sets of Network Information Data Mining. After completing the design of data mining rules based on association rules, when there is data mining information in the network environment, due to the large amount of information, there are many candidate sets for mining, which will increase the pressure of mining and cause the problem that the mining results cannot meet the expectations. Therefore, in order to effectively improve the mining efficiency of the network information data mining method based on association rules, it is necessary to filter its candidate set. According to the properties of candidate sets for network information data mining, assuming that T is the frequent x -item candidate set in data set P , the subset of all $x-1$ items of t can also be called the item set that makes it frequent $x-1$. Therefore, further analysis shows that T_x is the frequent x item candidate set in dataset P , so the number of $x-1$ item subsets included in the frequent $x-1$ candidate set L_{x-1} must be X . If an element will become an element in an x -dimensional frequent item set during mining, the number of occurrences of the element in the frequent $x-1$ item set must not be less than $x-1$. According to the above analysis, the candidate sets of network information data mining are screened. According to the properties of the candidate sets, this paper proposes an algorithm for further screening the number of candidate sets as follows: the L_{x-1} is trimmed before the CX generated by L_{x-1} . Count the actual times of all item arcs in L_{x-1} , delete the item set of items whose occurrence times are less than $x-1$ in L_{x-1} , and get L'_{x-1} . In order to distinguish the two, the above process is called clipping A, that is, clipping before candidate set screening. Then, it is called "pruning B" by using the pruning method provided by the association rule itself, that is, the pruning after candidate set filtering. Therefore, for a candidate set that needs to be mined, the screening results can be generated by the following algorithm: first, cut the candidate set a; use L_{x-1} to connect

one of the execution links to get the potential frequent item set in the candidate set; perform pruning B on the item set, and the final result is the candidate set of network information data mining after filtering [18].

3.4.3. Data Mining of Candidate Set Information. After completing the screening of candidate sets for network information data mining, data mining is carried out on the information in the candidate sets. Because the candidate sets still contain a large amount of data information, in the process of mining, this paper takes the programming idea as the basis combined with the data mining rules proposed in this paper, reconstructs the massive data of the candidate sets in the network environment, and uniformly classifies their texts. Calculate the occurrence probability of features under each category of the candidate set in the network environment. In the actual mining process, if the data under a certain feature appear frequently, the application value of mining will be reduced, resulting in the percentage of the mining data set in the important data set. Therefore, in order to effectively avoid this problem, this paper introduces another Apriori algorithm to reasonably allocate the weight set of each candidate set in the network environment when actually implementing association rules for data mining of network information candidate sets. The formula can be expressed as follows:

$$M = \lambda_i \times \log\left(\frac{Q}{d\lambda_i}\right). \tag{1}$$

In formula (1), M represents the weight allocation value of each candidate set in the network environment; Q represents the occurrence times of the candidate set in the network environment; D represents Apriori algorithm coefficient [19].

According to the above formula (1), complete the weight distribution of candidate sets, and on this basis, classify all candidate sets in the network environment so as to ensure

the accuracy of the final mining results and further improve the application significance of association rules. Through the above weight allocation results, the value obtained can be regarded as the evaluation result of the candidate set, and the network information data mining is completed by judging whether the final value of the evaluation output data is consistent with the value of the global cluster center. If the result shows that there is consistency between the two, it is considered that the value has certain application value. If the calculation result shows that there is no consistency between the two, it can be filtered by performing intelligent filtering, and the filtered data can be regarded as redundant data until all outliers in the network environment are mined [20].

4. Results and Discussion

In order to further verify the performance of the network information data mining method based on association rules proposed in this paper in practical application, the following comparative experiments are established: Taking the classical data set as the experimental sample, all association rules in the data set are mined, and the data sets are labeled from 1 to 9, respectively. Different labels correspond to different transactions, in which label 1 refers to transactions A, B, and E; Label 2 refers to transactions B and D; Label 3 refers to transactions B and C; Label 4 refers to transactions A, B, and D; Label 5 refers to transactions A and C; Label 6 refers to transactions B and C; Label 7 refers to transactions A and C; Label 8 refers to transactions A, B, C, and E; Label 9 refers to transactions A, B, and C. When there is a positive correlation between items in the current data set, it is considered that the improvement degree exceeds 1. When there is a negative correlation between items, it is considered that the degree of improvement is less than 1. Set the support level of the experimental environment to 0.3 and the confidence level to 0.8. Use the development tool of python3.1 to complete the application and implementation of the two mining methods through programming. After comparing the two mining methods, the percentage of the data set in the important data set is obtained.

In Figure 6, the p value represents the percentage of the data set obtained by this method or the traditional method in the important data set after mining. The larger the p value, the stronger the mining effectiveness of this method; on the contrary, the smaller the p value, the weaker the mining effectiveness of the method. The p value of this method is above 90.0%, while that of the traditional method is only within the range of 30.0%~70.0%, which is obviously higher. From the labels 1, 4, 8, and 9 in Figure 6, it can be seen that the effectiveness of the traditional method is worse when mining data sets with more transactions, while the method in this paper will not be affected by the number of transactions in the data set in the process of mining data sets. Therefore, through comparative experiments, it is further proved that the network information data mining method based on association rules proposed in this paper is more effective in practical applications. It can complete the information data mining with higher utilization value and improve the effective utilization of data.

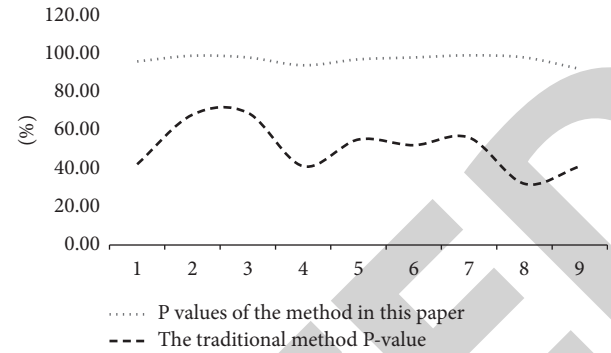


FIGURE 6: Comparison of experimental results of two mining methods.

5. Conclusion

The explosion of the information age has led to the rapid increase of various data resources. However, the difference between the increase of data and the lag of data analysis is also growing. Most researchers hope to mine the deep value of data through scientific means, so data mining has become the mainstream technology to solve data analysis problems. It makes up for the shortcomings of traditional analysis methods and scientifically processes data. Only when effective knowledge information hidden in data is discovered in time can it further serve human development, and data resources can be truly utilized, which also means the real arrival of the era of big data. The main steps of the method in network information data mining are as follows: (1) designing data mining rules based on association rules; (2) screening candidate sets of network information data mining; (3) candidate set information data mining. Through comparative experiments, it is further proved that the network information data mining method based on association rules proposed in this paper is more effective in practical applications, can complete the information data mining with a higher utilization value, and improve the effective utilization of data.

Data Availability

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

Conflicts of Interest

The author declares that they have no conflicts of interest.

References

- [1] I. Kouveliotis-Lysikatos, N. Hatzigiorgiou, Y. Liu, and F. Wu, "Towards an ipg," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 1, pp. 1–11, 2022.
- [2] M. Bhatia and A. Rana, "A mathematical approach to optimize crop allocation – a linear programming model," *International Journal of Design & Nature and Ecodynamics*, vol. 15, no. 2, pp. 245–252, 2020.

- [3] S. Pu, W. Shi, J. Xu, and A. Nedić, "Push-pull gradient methods for distributed optimization in networks," *IEEE Transactions on Automatic Control*, vol. 66, no. 1, pp. 1–16, 2020.
- [4] P. Hein, M. Bernhagen, and A. Bullinger-Hoffmann, "Two is better than one. Improved attention guiding in AR by combining techniques," *IEEE Computer Graphics and Applications*, vol. 40, no. 5, pp. 57–66, 2020.
- [5] A. A. Jalal and B. H. Ali, "Text documents clustering using data mining techniques," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 1, pp. 664–670, 2021.
- [6] S. H. Lee, B. S. Lin, H. C. Lee et al., "Artificial intelligence based assessment system for evaluating suitable range of heel height," *IEEE Access*, vol. 9, pp. 38374–38385, 2021.
- [7] H. Hockx-Yu, "Invisible Women, Data Bias in A World Designed For Men: Caroline Cirado Perez," vol. 5, no. 1, pp. 1–5, Abrams Press, New York, 2019.
- [8] Y. Jeong, M. Hwang, and W. Sung, "Autoencoder-based learning contribution measurement method for training data selection," *Journal of KIISE*, vol. 48, no. 2, pp. 195–200, 2021.
- [9] H. Chen, W. Wang, G. Li, and Y. Shi, "A quaternion-embedded capsule network model for knowledge graph completion," *IEEE Access*, vol. 8, pp. 100890–100904, 2020.
- [10] R. Dadsena, D. Sadukhan, and R. Swaminathan, "Rotational moment shape feature extraction and decision tree based discrimination of mild cognitive impairment conditions using MR image processing," *Biomedical Sciences Instrumentation*, vol. 57, no. 2, pp. 228–233, 2021.
- [11] R. Duo, X. Nie, N. Yang, C. Yue, and Y. Wang, "Anomaly detection and attack classification for train real-time ethernet," *IEEE Access*, vol. 9, pp. 22528–22541, 2021.
- [12] R. Mondal, A. Chakraborty, J. Dey, and S. Halder, "Optimal fractional order $PI^\lambda D^\mu$ controller for stabilization of cart-inverted pendulum system: experimental results," *Asian Journal of Control*, vol. 22, no. 3, pp. 1345–1359, 2020.
- [13] Y. Wang, "Prediction of Rockburst Risk in Coal Mines Based on a Locally Weighted C4.5 Algorithm," *IEEE Access*, vol. 9, pp. 15149–15155, 2021.
- [14] S. A. El-Khatib, Y. A. Skobtsov, and S. I. Rodzin, "Comparison of Hybrid ACO-k-means algorithm and Grub cut for MRI images segmentation," *Procedia Computer Science*, vol. 186, no. 11, pp. 316–322, 2021.
- [15] H. Xie, "Research and case analysis of Apriori algorithm based on mining frequent item-sets," *Open Journal of Social Sciences*, vol. 09, no. 04, pp. 458–468, 2021.
- [16] G. Dhiman, V. Vinoth Kumar, A. Kaur, and A. Sharma, "Don: deep learning and optimization-based framework for detection of novel coronavirus disease using x-ray images," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 13, no. 2, pp. 260–272, 2021.
- [17] M. Bradha, N. Balakrishnan, S. Suvi et al., "Experimental, Computational Analysis of Butein and Lanceoletin for Natural Dye-Sensitized Solar Cells and Stabilizing Efficiency by IoT," *Environment, Development and Sustainability*, vol. 24, no. 7, 2021.
- [18] Q. Liu, N. Cheng, Q. He et al., "Meteorological conditions and their effects on the relationship between aerosol optical depth and macro-physical properties of warm clouds over shanghai based on modis," *Atmospheric Pollution Research*, vol. 11, no. 9, pp. 1637–1644, 2020.
- [19] P. Ajay, B. Nagaraj, R. A. Kumar, R. Huang, and P. Ananthi, "Unsupervised hyperspectral microscopic image segmentation using deep embedded clustering algorithm," *Scanning*, vol. 2022, Article ID 1200860, 9 pages, 2022.
- [20] G. Veselov, A. Tselykh, A. Sharma, and R. Huang, "Special issue on applications of artificial intelligence in evolution of smart cities and societies," *Informatica*, vol. 45, no. 5, p. 603, 2021, <http://www.informatica.si/index.php/informatica/article/view/3600>.