
Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 4, pp. 2220–2228, DOI 10.12694/scpe.v25i4.2941

APPLICATION OF DEEP LEARNING AND COMPUTER DATA MINING TECHNOLOGY IN ELECTRONIC INFORMATION ENGINEERING MANAGEMENT

KUN JIANG*

Abstract. This article studies the application of deep learning and computer data mining technology in electronic information engineering management to meet the library's demand for larger collection space and alleviate the management pressure of book preservation, borrowing, and return. This article also utilizes the general information mining function to further improve the information retrieval function. The conclusion of this article is as follows: Based on the traditional April algorithm, an improved address based April algorithm is proposed. The improved Apriori algorithm can reduce the final number of permanent data packets and save about 70% of time. For the minimum supported changes, the improved Apriori algorithm has lower execution time and approximately 60% time reduction. This article develops a data mining algorithm that is more suitable for electronic library management information systems based on existing data mining technologies. Has clear theoretical and practical significance.

Key words: Data mining, Cluster analysis, Apriori algorithm, Electronic Library

1. Introduction. The rapid development of computer technology, Internet technology, and information technology has brought about significant changes in people's production methods and lifestyles, as well as profound changes in the library industry. Libraries have long considered paper books to be the most important resource. However, as information technology deepens, paper books need more collection space to cover a more comprehensive amount of knowledge, which also puts great pressure on the management of book preservation, borrowing and return [1]. In this case, colleges and universities, scientific research institutions and other units have begun to build electronic libraries. Electronic library is a library model that stores information in the form of e-books. The emergence of electronic library benefits from the emergence of electronic production technology of books and publications [2]. As soon as the electronic library appears, it shows a variety of advantages over the traditional library, and quickly gets the favor of users.

Electronic libraries use disks and optical discs as storage, which have a much higher storage capacity than paper books, resulting in the storage of massive amounts of information; The borrowing of e-books is more convenient, and users can download them through the internet; E-books are easy to keep and extend their service life indefinitely: in addition to text information, e-books can also conveniently store sound information, image information and video information [3].

Since the emergence of electronic library, it has been proved by indisputable facts that it will become the mainstream model in the field of library. However, e-library also puts forward a series of new challenges to developers.

The most important problem is how to quickly and accurately retrieve what users need in the massive book information. Mass storage is an important advantage of e-library, but the larger the storage, the greater the difficulty for the retrieval process.

In this case, data mining has become the first choice to improve the retrieval performance and improve the service quality of E-library[4]. From a processing point of view, data mining technology is the process of extracting information and extracting knowledge using various algorithms. Some of these algorithms depend on statistics, some on artificial intelligence, and some on machine learning perspectives. No matter what type of algorithm is adopted, data mining is a technology with independent analysis ability and automatic retrieval of knowledge [5]. Even if the information provided by users is not comprehensive enough, data mining can still retrieve or extract useful data for users, which is the advantage of data mining. This is undoubtedly of

*Xinxiang Vocational and Technical College, Xinxiang, Henan 453006, China (Corresponding author)

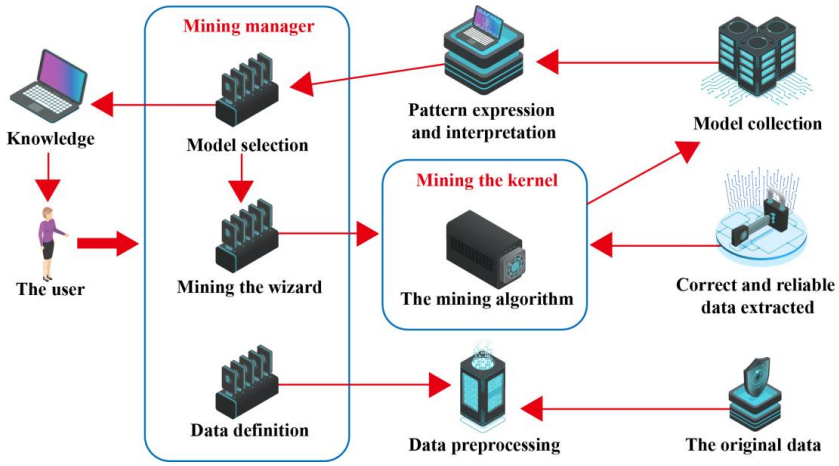

Fig. 1.1: Application of computer data mining technology in electronic information engineering

great significance for electronic libraries [6]. In recent years, data mining technology has been used in similar information management systems.

In the case of an electronic library system, the purpose of using data mining technology in it is, on the one hand, to obtain general information and, on the other hand, to complete the function of sorting further information. From this perspective, this article develops a data mining algorithm that is more suitable for electronic library management information systems based on existing data mining technologies, which has certain theoretical and practical significance [7]. Figure 1.1 shows the use of computer data extraction technology in electronic information engineering.

2. Literature Review. After the 1990s, human society officially entered the information society, and the amount of data showed an explosive growth trend. It has been difficult to effectively sort out and analyze such large-scale data by relying on human ability.

Data mining technology can quickly and accurately organize and analyze information, which has become an inevitable technical requirement [8]. After entering the 21st century, data mining technology has been recognized as the first of the key technologies that will have a significant impact on human production and life in the next decade, and has gradually become one of the key targets of enterprise investment. In this case, data mining technology has developed rapidly [9]. Attari, M. and others systematically summarized data mining technology. According to the development of data mining technology, data mining technology is positioned as a multi technology cross field of database technology, statistical theory, machine intelligence technology, visualization technology and digital technology [10]. Wang, R. and others pointed out that the nature of data mining technology determines that it is inseparable from intelligent algorithms. Neural network intelligent algorithms, fuzzy theory intelligent algorithms and knowledge theory intelligent algorithms are the basic algorithms of data mining technology. Mining technology is a knowledge integration tool. Data mining technology can use many existing knowledge, such as information retrieval knowledge, signal processing knowledge, image processing knowledge, web database knowledge, pattern recognition knowledge, etc. [11]. Gao, J. and others proposed a data mining method based on maximum range search, and constructed a visual data mining system according to this method. This system can not only automatically execute the data mining process, but also be very convenient for human operation [12]. Parvez, I. and others focused on the data mining technology with sequence attributes on the basis of general data mining technology, and proposed a data mining method based on association rules, which is specially used for data mining of sequence data. This is a very important problem in the process of data mining [13]. Only the accurate segmentation of the information to be extracted before mining is conducive to the accurate positioning of the extraction results. With the development of technology, a data mining method based on gene distribution and cluster analysis is proposed, which can deal with more

subtle data mining problems. Li, J et al. pointed out that rough set theory is a very effective data mining tool that can effectively preserve important attribute degrees and exclude non important attribute degrees. We can use rough set theory to achieve quantitative mining. For this type of data mining method, others have developed a set of evaluation indicators that can evaluate the mining performance of different methods [14].

3. Research methods.

3.1. Data mining. The key of the concept of data mining is mining two words, which determines that it is different from simple data processing and data integration. Data mining discovers more valuable information hidden behind surface data information, such as the development and change trends of certain data, typical feature attributes of certain data, and correlations between certain data.

In terms of implementation, data mining must be combined with database theory, machine recognition theory and intelligent learning theory to achieve better results [15-16]. From the general implementation process of data processing, building a data warehouse is a prerequisite work. Only when the data is stored in the data warehouse, can data mining have operable objects. Therefore, data warehouse is the basis of data mining. Data mining is to find or extract more valuable information in the data warehouse [17].

OLAP technology also comes from the semantic translation of loanwords. Its corresponding English word is online analytical process, and its Chinese meaning is online analytical processing. Data mining technology is based on a series of assumptions, but these assumptions cannot be verified by data mining technology itself. OLAP technology just solves this problem. It can verify whether the results of data mining meet the assumptions [18].

Machine learning refers to that after some intelligent algorithms are executed on computers and other machines, they can learn some knowledge and form memory, so as to deal with subsequent problems more effectively. Data mining technology sometimes needs to use machine learning methods. For example, in large medical databases, patients have a lot of information and different conditions. At the same time, these information will also have the same or similar situations.

At this time, the implementation efficiency of data mining technology can be improved by using machine learning [19]. The core task of data mining is to select the most suitable mining algorithm. Due to the many existing mining algorithms, different algorithms have different application ranges. Choosing the most suitable and accurate algorithm for one's own problems directly determines whether the mining performance can achieve the expected results [20].

3.2. Cluster analysis. In the automatic service of e-library, the main service object is readers who need various types of books. The final service result is to provide them with e-books they may need for downloading. Therefore, if we can classify reader data and book data appropriately before deep mining, we will complete the service process faster and more accurately. In this paper, K-Mean is an average clustering method that analyzes important e-library data into clusters, so that the subsequent mining process can be better realized. Cluster analysis is the process of further subdividing an original data set according to abstract attributes or physical attributes to form multiple subclasses or sub clusters. This refinement can make the subsequent mining objects more clear, the scope more narrow and the results more accurate.

The K-Means cluster analysis algorithm is a common way to create a distance-based cluster. It can realize completely automatic and unattended cluster analysis, and the whole implementation process is also very simple. Especially when the number of clusters is determined, the clustering effect of this method is very obvious. The basic idea of K-means clustering analysis is: firstly, k data positions are randomly determined as the central position of cluster analysis, and then the distances from each point in the original data set to these central positions are calculated and included in each cluster.Then, based on the new data of each cluster, we need to continuously update the position of the cluster center and repeat the clustering, clustering again until the position of each cluster center remains unchanged, and then the entire clustering process ends [21,22].

As a clustering method based on distance, K-means clustering considers that the closer the distance between two data is, the higher the similarity is. If the distance between two data is, the lower the similarity is. After this judgment and processing, K-means clustering will get a very compact and independent cluster.

The implementation process of K-means clustering is as follows:

1. Randomly select one data from all the data as the central position to perform clustering.

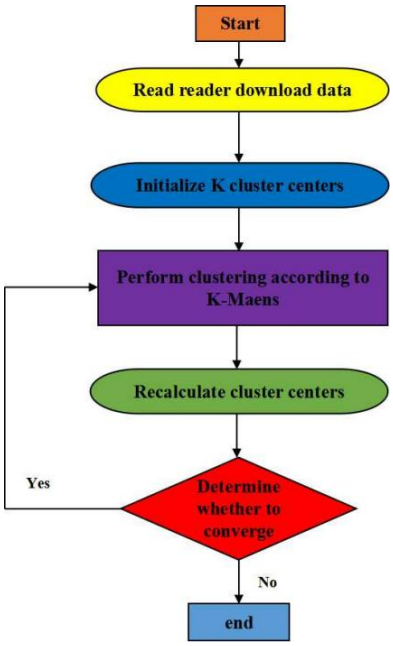

Fig. 3.1: Reader cluster analysis

2. For other data, the distance from the center of the cluster is calculated, and each data is classified according to the results of the calculation.
3. According to the data that has been summarized into each cluster, calculate the average value of each data and update the position of the cluster center.
4. According to the new clustering center, repeat the work from step 2 to step 3 until the clustering results meet the preset conditions, and the clustering ends.

K-means clustering selects a cluster center from the original data set. These clusters are shown in Equation 3.1:

$$C = \{C_k | i = 1, 2, ..., K\} \tag{3.1}$$

If the centers of these clusters are, the distance between other data and these centers can be calculated by the following formula, such as Equation 3.2:

$$J(c_k) = \sum_{x_i \in C_k} ||x_i - \mu_k||^2 \tag{3.2}$$

In the formula, x_i represents each data.

To determine whether the final clustering is completed, it is necessary to determine whether the sum of squares of all distances is the smallest, as shown in Equation 3.3:

$$J(C) = \sum_{k=1}^{K} J(c_k) \tag{3.3}$$

When reaches the minimum, the clustering is determined as the end. In this paper, K-means clustering analysis is performed on users based on the number of downloads of readers [23]. The specific process see Figure 3.1.

In the electronic library, in addition to user data, book data is another kind of very important data. If we can effectively cluster the download frequency of books, it will help to analyze what are hot books, also help to

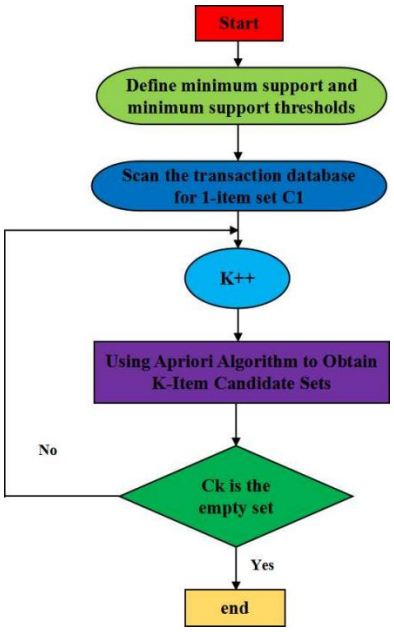

Fig. 3.2: Apriori algorithm flow

analyze the interests of users, and is more conducive to the rational purchase of subsequent e-book resources [24].

3.3. Apriori algorithm. Since the advent of the a priori algorithm, it has always been the main algorithm in the field of data mining and has been widely used in various data mining systems. Its core idea is to perform frequent item search on the database to be processed, and establish a set of frequent items to form a priori knowledge for subsequent use. Then, we determine the association rules of the original dataset by gradually searching for the relationships between frequent items in the database system, providing decision-making basis for users to deeply analyze the data.

Frequent item sets have two properties, which become the theoretical basis for the construction of Apriori algorithm [25]. These two properties are:

First, for infrequent item sets, its parent set must also be infrequent item sets.

Second, for frequent item sets, its descendant sets must also be frequent item sets.

According to these two properties, if a set can be represented as an infrequent N-term pattern, then any n + 1-term pattern to represent the set is also infrequent. Accordingly, the Apriori algorithm is formed, as shown in Figure 3.2.

3.4. Improved Apriori algorithm. This article discusses some of the improvement measures to address the common problems of a priori algorithms, to reduce the complexity of frequency packages, and to improve the efficiency of algorithm performance. The basic strategy for improving the algorithm is to increase the address configuration to increase the pruning efficiency of the Apriori algorithm. A priori algorithm is a very classic data mining algorithm. It is widely used because of its simple theory, convenient operation and high mining efficiency. Of course, Apriori algorithm is not absolutely perfect. It also has some typical problems, as follows:

1. The candidate frequent set CK formed by the algorithm is too large.
   In fact, among the data objects processed by Apriori algorithm, the more data items and entry items, the scale of candidate frequent sets will increase exponentially. This will greatly increase the processing time of data mining process, greatly increase the occupation of memory space, and greatly reduce the execution speed of the whole Apriori algorithm.

Table 4.1: Clustering results of some books

Book number	Book name	Average number of downloads per year	Total Downloads	Book type
0A041138	Three days proficient in office	51	106	Normal books
0Z012153	Yi Zhongtian's view of the Three Kingdoms	220	412	Hot books
0C080920	CET4 real exercise	41	83	Normal books
0Z012275	Political situation of the two Song Dynasties	112	170	Hot books
0D110003	Biological basis	15	21	Unpopular books

2. The algorithm performs too many scans on the processing object.
   The key is to scan the candidate items frequently and execute the Apriori algorithm frequently. If the final frequent item set is k-level, the scanning of data objects must reach k-1 times. When the data scale increases, the corresponding scanning times will be very many. Such multiple scans will be directly congested by the I / O of the computer, greatly reducing the execution efficiency of the algorithm.
3. Many association rules formed by the algorithm are redundant.
   As the scale of processing objects continues to expand, the number of candidate frequent set operations, self linking operations, and scans will significantly increase, and the final association rules will also rapidly expand. However, not all association rules have guiding significance for practical problems. Many association rules are invalid and redundant. When there are too many association rules, it is difficult for users to judge which rules can be used.

4. Result Analysis.

4.1. Cluster analysis examples. According to the different classification basis of cluster analysis, the clustering results of some books are obtained, as shown in Table 4.1. According to this classification result, the download types of different books are obtained. Among them, the book "Biological basis" is an unpopular book; Proficient in office in three days "Four true exercises" these two books are normal books; "Yi Zhongtian's view of the Three Kingdoms" and "Political situation of the two Song Dynasties" are two hot books. It shows that the clustering results have good results, which can be used for the next experiment.

4.2. Performance test of improved Apriori algorithm. Performance tests of the two algorithms are performed to better understand the performance of the a priori algorithm and the improved Apriori algorithm. From the perspective of distance, the improved Apriori algorithm does not scan the original data set many times, but uses one scan and address traversal to complete the search of frequent items, which greatly saves time and cost. In addition, judging only frequent items from address comparison can not only realize effective search, but also greatly improve the search efficiency. Address based storage also greatly saves memory space. Based on these characteristics, the improved Apriori algorithm greatly improves the mining efficiency.

The following experimental methods are used in this paper to quantify the performance of the improved Apriori algorithm. The specific test process is divided into two methods: the first method is to fix the minimum support rating in the Apriori algorithm and the improved Apriori algorithm, and the two algorithms check the time to generate the last frequency items of the same series, respectively. The second method is to convert the minimum support rating in the Apriori algorithm, improve the Apriori algorithm, and check the execution times of two algorithms with different lower support levels. Minimum support does not change Set the minimum support to 4, the set of candidate frequent and complex items is shown in Table 4.1, and the time of the frequent item set finally generated by the two algorithms is shown in Table 4.3.

It can be seen from table 4.3 that in the mining process of the improved Apriori algorithm, the number of final frequent item sets becomes less and the time is less, saving about 70%. Draw the data in Table 4.3 into

Table 4.2: Configuration of candidate frequency complex item set

Candidate frequency items	Set C_1	Set C_2	Set C_3	Set C_4	Set C_5	Set C_6
Number of frequent items	21	345	1077	646	64	7

Table 4.3: Time of generating the final frequent item set by the two algorithms

Algorithm	L1	L2	L3	L4	L5	L6	Total time
Apriori algorithm	135	160	237	181	126	110	1001
Improved Apriori algorithm	136	78	100	62	20	-2	370

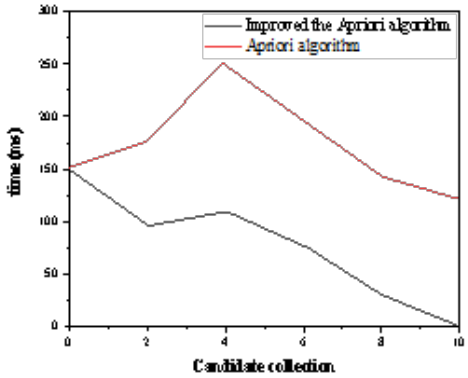

Fig. 4.1: Execution time of two algorithms

the following curve form, as shown in Figure 4.1.

As can be seen in Figure 4.1, the execution time of the improved Apriori algorithm is significantly reduced under the established minimum support.

As shown in Figure 4.2, the degree of minimum support, the minimum degree of support for multiple conversions, and the change in performance time of the Apriori algorithm and the improved Apriori algorithm were examined.

5. Conclusion. In recent years, data mining technology has been applied in similar information management systems. From a processing point of view, data mining technology is the process of extracting information and extracting knowledge using various algorithms. Some of these algorithms are considered from a statistical point of view and some from an artificial intelligence point of view, and some from the perspective of machine learning. No matter what type of algorithm is used, data mining is a technology with independent analysis ability and automatic retrieval of knowledge. Even if the information provided by users is not comprehensive enough, data mining can still retrieve or extract useful data for users, which is the advantage of data mining. This is undoubtedly of great significance for electronic libraries. For electronic library systems, data mining technology can not only complete general information retrieval functions, but also further improve information sorting functions. From this point of view, based on the existing data mining technology, this paper designs a data mining algorithm more in line with the electronic library management information system, which has certain theoretical significance and practical value for improving the service quality of electronic library and expanding the application of data mining algorithm.

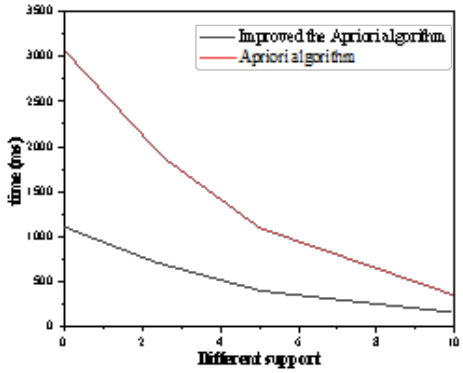

Fig. 4.2: Execution time of two algorithms

This paper has done some work in the following aspects:

1. Based on the principle of K-means clustering, this paper designs the clustering analysis function for the electronic library management information system, and arranges the data from two levels: user data and book data.
2. Based on the traditional Apriori algorithm, an improved address-based Apriori algorithm was developed. This method has been experimentally proven to significantly improve the speed of the cutting operation in the data mining process, improve the efficiency of the data mining system.

REFERENCES

[1] Farhadi, Esmaeily, A., & Najafzadeh, M. . (2021). Developing a decision tree based on data mining method for detecting the influential parameters on the power of flood destruction. Amirkabir (Journal of Science and Technology), 53(5), 5-5.

[2] Zhou, R., Chen, H., Chen, H., Liu, E., & Jiang, S. . (2021). Research on traffic situation analysis for urban road network through spatiotemporal data mining: a case study of xi'an, china. IEEE Access, PP(99), 1-1.

[3] Moayedi, H., Abdullahi, M. M., Nguyen, H., & Rashid, A. . (2021). Comparison of dragonfly algorithm and harris hawks optimization evolutionary data mining techniques for the assessment of bearing capacity of footings over two-layer foundation soils. Engineering with Computers, 37(1), 437-447.

[4] Yang, T., Zhang, L., Kim, T., Hong, Y., & Peng, Q. . (2021). A large-scale comparison of artificial intelligence and data mining (ai&dm) techniques in simulating reservoir releases over the upper colorado region. Journal of Hydrology, 602(6), 126723.

[5] Booker, E. P., & Jabbour, G. E. . (2021). Antiviral nanoparticle ligands identified with datamining and high-throughput virtual screening. RSC Adv., 11(37), 23136-23143.

[6] Li, Y., Shyamasundar, R. K., & Wang, X. . (2021). Special issue on computational intelligence for social media data mining and knowledge discovery. Computational Intelligence, 37(2), 658-659.

[7] Liu, Y., Yu, Z., & Yang, Y. . (2021). Diabetes risk data mining method based on electronic medical record analysis. Journal of Healthcare Engineering, 2021(6), 1-11.

[8] Wu, B., Qin, D., Hu, J., & Liu, Y. . (2021). Experimental data mining research on factors influencing friction coefficient of wet clutch. Journal of Tribology, 143(12), 1-14.

[9] Bardak, S., Bardak, T., Peker, H., Szen, E., & Abuk, Y. . (2021). Predicting effects of selected impregnation processes on the observed bending strength of wood, with use of data mining models. Bioresources, 16(3), 4891-4904.

[10] Attari, M., Ejlaly, B., Heidarpour, H., & Ala, A. . (2021). Application of data mining techniques for the investigation of factors affecting transportation enterprises. IEEE Transactions on Intelligent Transportation Systems, PP(99), 1-16.

[11] Wang, R. . (2021). Exploration of data mining algorithms of an online learning behaviour log based on cloud computing. International Journal of Continuing Engineering Education and Life-Long Learning, 31(3), 371.

[12] Gao, J., Yue, X. G., Hao, L., Crabbe, M., & Duarte, N. . (2021). Optimization analysis and implementation of online wisdom teaching mode in cloud classroom based on data mining and processing. International Journal of Emerging Technologies in Learning (iJET), 16(1), 205.

[13] Parvez, I. . (2021). Generation of hydro energy by using data mining algorithm for cascaded hydropower plant. Energies, 14(2), 298.

[14] Li, J. . (2021). Analysis of the mental health of urban migrant children based on cloud computing and data mining algorithm models. Scientific Programming, 2021(5), 1-7.
[15] Gu, J. . (2021). An effectiveness model of vocational education mode reform based on data mining. International Journal of Continuing Engineering Education and Life-Long Learning, 31(1), 1.
[16] Chen, P., & Yu, L. . (2021). Use of data mining technologies in an english online test results management system. International Journal of Emerging Technologies in Learning (iJET), 16(9), 166.
[17] Tao, L. . (2021). Application of data mining in the analysis of martial arts athlete competition skills and tactics. Journal of Healthcare Engineering, 2021(4), 1-6.
[18] Wang, B. . (2021). Multimedia filtering analysis of massive information combined with data mining algorithms. Advances in Multimedia, 2021(3), 1-7.
[19] Wang, C., Bi, J., Sai, Q., & Yuan, Z. . (2021). Analysis and prediction of carsharing demand based on data mining methods. Algorithms, 14(6), 179.
[20] Druskin, V.,Mamonov, & Zaslavsky, M. . (2022). Distance preserving model order reduction of graph-laplacians and cluster analysis. Journal of Scientific Computing, 90(1), 1-30.
[21] Limin, Pan, X., Qin, S., & Luo. (2020). Dsp-tmm: a robust cluster analysis method based on diversity self-paced t-mixture model. Journal of Beijing Institute of Technology, v.29;No.106(04), 100-112.
[22] Hu, J., Xing, Y., Han, M., Wang, F., Zhao, K., & Che, X. . (2022). Nonnegative matrix tri-factorization based clustering in a heterogeneous information network with star network schema. Tsinghua Science and Technology, 27(2), 386-395.
[23] Hong, J., Tamakloe, R., & Park, D. . (2020). Discovering insightful rules among truck crash characteristics using apriori algorithm. Journal of Advanced Transportation, 2020(2), 1-16.
[24] Tian, M., Zhang, L., Guo, P., Zhang, H., & Xue, A. . (2020). Data dependence analysis for defects data of relay protection devices based on apriori algorithm. IEEE Access, PP(99), 1-1.
[25] Cai, Q. . (2020). Cause analysis of traffic accidents on urban roads based on an improved association rule mining algorithm. IEEE Access, PP(99), 1-1.

Edited by: B. Nagaraj M.E.
Special issue on: Deep Learning-Based Advanced Research Trends in Scalable Computing
Received: Nov 30, 2023
Accepted: Jan 24, 2024