Research Article

Data Mining Technology in Book Copyright Information Management Decision System

Song Lifang

Labor Union Renmin University of China, Beijing 100872, China

Correspondence should be addressed to Song Lifang; slifang@ruc.edu.cn

Received 9 August 2021; Revised 10 September 2021; Accepted 13 September 2021; Published 30 September 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Song Lifang. 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.

Today is an era of data “big bang”; Internet information technology is widely used in various fields of society. As an indispensable spiritual food in people’s daily life, books are increasing in number and scale. In order to better manage book information, people have introduced data mining technology. Based on this, this article takes the research and application of data mining technology in book copyright information management decision-making system as the theme, explores the role of data mining technology in book copyright information management, and aims to provide reference for our country’s book copyright information management and decision-making. This article first introduces the common algorithms of data mining technology and then elaborates on the advantages and effectiveness of the association rule method in data mining. Aiming at some defects of the original Apriori algorithm of the association rule method, an improved Apriori algorithm is proposed. After taking the library book information management system and database of a university in our province as the experimental research object, the performance gap between the two algorithms is compared through experiments, and it is concluded that when the number of transaction set item records is less than 1400, the Apriori algorithm performs better, and when the number of records in the transaction set is greater than 1400, the improved Apriori algorithm is obviously more advantageous. The research results show that the introduction and application of data mining technology make the information management of books more efficient and convenient, and it is more convenient for the management and decision-making of book copyright information.

1. Introduction

1.1. Background and Significance. Since the beginning of the 21st century, our country’s science and technology have developed rapidly, and Internet technology has been more and more widely used in all walks of life and people’s daily lives. With the advent of the information age, people are becoming more and more adept at using high-tech means to collect data and deal with problems and are most widely used in economics and trade, scientific research, project management, etc. The society is constantly developing, followed by the continuous advancement of information technology. In order to adapt to the requirements of the times, various information technology processing methods have emerged, and data mining technology is one of them.

Books are the spiritual food for people to acquire knowledge and increase their knowledge. With the

continuous expansion of social demand, the storage and scale of various books and materials in the library are expanding, which has brought tremendous pressure to the management of books. The problem of book copyright in library management has always been paid attention to. Traditional manual management not only has low work efficiency and high error rate, but the heavy workload also consumes huge amounts of human resources. However, the use of data mining technology can realize data mining and cluster analysis of massive book information and achieve the purpose of efficient, fast, and intelligent management of book information. Data mining technology has the powerful function of quickly processing and analyzing information [1]. Through the reading and analysis of information, an information database will be built within the system, and clustering will be made according to the different characteristics of the acquired information, which greatly improves

book management efficiency and quality. It can be seen that the application of data mining technology in book management is bound to be the general trend.

1.2. Related Work. Since data mining technology has been widely used in book and library management, many experts and scholars have conducted research on this aspect. Ye thinks that how to quickly retrieve the required information from the vast data information requires Web data mining technology [2]. Its algorithm and its application in information management are topics worthy of study. Nowadays, Web data mining plays an important role in information management fields such as scientific literature retrieval. Tian also said that the library will involve intellectual property issues in the provision of various services for readers and its own management, especially under the conditions of traditional services and informatization [3]. There are many intellectual property issues involved in services and management, and the characteristics are extremely complex, which requires a powerful information technology means to process and analyze it. In addition, Abu Sirhan and Abdrabbo also pointed out some methods of using digital rights management (DRM) technology to protect digital libraries, especially the methods applied in the digital library of the University of Jordan [4]. They said that at the digital library of the University of Jordan, the degree of DRM has been used. Most universities in Jordan use different protection methods, including encoding, identification, authenticity, and digital watermarking. However, Jordan University has not adopted methods including digital signatures, digital fingerprints, photocopy detection systems, and payment systems. The use of the copyright management system is restricted, which indicates that there are still certain weaknesses in the use of technical protection in the libraries surveyed. In short, there are many researches on the application of data mining technology in book information management. This article proposes a new and improved Apriori algorithm based on the research results of the predecessors, which aims to provide data mining technology in the field of book information management [5]. Wu believes that big data is widely regarded as one of the most powerful driving forces to promote productivity, improve efficiency, and support innovation. Exploring the power of big data and turning big data into big value is highly anticipated. In order to answer the interesting question of whether there is an inherent correlation between the two trends of big data and green challenges, his recent study investigated the issue of greening the entire life cycle of big data systems. He hopes to discover the relationship between the trend of the big data era and the trend of the new generation of green revolution through a comprehensive and panoramic literature survey on various green goals of big data technology, as well as discussions on related challenges and future dilemmas [6].

1.3. Innovations in This Article. Based on the introduction and research of data warehouse, online analytical processing, and data mining theory and related technologies, this paper

analyzes the feasibility of applying data warehouse technology, online analytical processing technology, and data mining technology to decision support systems in university libraries. Design the data warehouse of the university library, and realize the extraction, conversion, and loading of the data of the university library into the data warehouse. The innovations of this paper are mainly reflected in the following aspects: (1) it introduces the common algorithms of data mining technology in detail and proposes an improved Apriori algorithm for the defects of the original Apriori algorithm in the association rule method. The algorithm shown in the calculation of a large number and complex transaction set data shows better performance; (2) it takes the book information management system and database of a university library in our province as an example, so that the content of the article research has a carrier and is more popular. It is easy to understand and more vivid; and (3) the theme of this article is closely related to the background of the times and has great social practical significance and discussion value.

2. Application Strategy of Data Mining Technology in the Book Copyright Information Management Decision System

2.1. Data Mining Technology

2.1.1. Overview of Data Mining. Data mining first appeared in the 1990s. At first, it was just a database system for managing simple data. Later, with the continuous development and progress of society, the database system slowly developed into the management of images, graphics, video, audio, and electronics generated by computers. Files and other complex data information are also included [7]. It can be said that data mining is a derivative of the rapid development and continuous updating of science and technology in the new era. Data mining belongs to the field of computer science [8, 9]. It is a process of finding out the laws and information hidden behind the data through algorithms from a large amount of data. In recent years, the advantages of data mining in processing data and information have been widely recognized, which has attracted great attention from the information industry, especially in the fields of economics and trade, scientific research, and resource information management [10]. Its huge database lacks effective means. Analysis and processing have become a useless empty shell. They urgently need to convert these data into useful information and knowledge to help improve the efficiency and quality of their work management and the rate of industry competition [11].

2.1.2. Data Mining Objects. Data mining has a wide range of objects, which can be digital or nonmathematical. It includes any type of data source [12]: structured, such as relational databases; semistructured and heterogeneous, such as data warehouses, text, multimedia video images, spatial data, financial time series data, Web data, etc. [13, 14]. Through the mining and analysis of these data, we can find out the hidden rules, relationships, and knowledge behind it and apply

them to information management, query optimization, decision support, and maintenance of the data itself [15, 16].

2.1.3. Steps of Data Mining

Step 1. Understand the data and clearly define it. Before conducting data mining and knowledge discovery on a database, you must first understand the data information and business problems of the database, clarify the definition and requirements of the target, and conduct targeted mining and analysis, so as to achieve the best mining effect. The knowledge and information excavated is also the most accurate and effective [17].

Step 2. Build a data mining library. After having a preliminary understanding of the original database, collect and organize effective data information [18]; then describe it, perform data quality assessment and data cleaning, and finally integrate and construct metadata to establish its own data mining library, and continue to load and maintain data mining library [19, 20].

Step 3. Process and analyze data. After the data mining library is established, the data must be processed and analyzed [21]. The purpose of the analysis is to find the data fields that have the greatest impact on the expected output. Usually, for huge databases, we will use certain statistical analysis software to assist in completing the data analysis work, such as SPSS series statistical analysis software [22].

Step 4. Prepare data. According to the analyzed data, select valid variables as useful data for establishing the model. This process involves four parts: variable selection, record selection, new variable creation, and variable conversion [23].

Step 5. Establish a data mining model. When building a model, various factors should be considered comprehensively. The selected variables should be divided into training group and test group. The data of the training group is used to build the model. After the model is established, the test group data should be used to test the accuracy of the model. The optimization of the effect after verification is the modeling success [24, 25].

Step 6. Evaluate the model. The establishment of many models is based on assumptions, and there will be many gaps in actual application. Therefore, when evaluating models, you should first select applications in a small range. After the effects are significant, they can be promoted on a large scale. From a commercial perspective, industry experts verify the correctness of the data mining results.

Step 7. Implement the application. After the model is established and verified, it can be specifically applied to examples for data mining.

2.1.4. Common Algorithms of Data Mining Technology

(i) *Neural Network Algorithm.* Neural network algorithm is an artificial intelligence calculation method based on computer science that imitates the structure and function of human brain neurons. It has strong self-learning ability and feedback characteristics and

the ability to wirelessly approximate nonlinear functions. It can acquire knowledge through learning and training and apply knowledge and solve problems. It is a nonlinear prediction model with excellent performance. Usually the learning method of neural network is reflected in the modification and adjustment of weights. It is divided into two forms: feedforward and feedback. Common algorithms include BP neural network algorithm, radial basis neural network algorithm, and recurrent neural network algorithm. The advantage is that it is anti-interferential and can make accurate predictions for complex situations, but the disadvantage is that it is not suitable for high-dimensional predictions and often requires dimensionality reduction.

- (ii) *Genetic Algorithm.* The genetic algorithm simulates the principle of survival of the fittest in nature and adopts a biological evolution theory that combines biological heredity and mutation, natural selection, and elimination as a machine learning method to implement rules. It has implicit parallelism and is easy to combine with other models. The advantage is that multiple types of data can be processed in parallel at the same time, but the disadvantage is that the amount of calculation is generally large and contains too many parameters, which is difficult to implement.
- (iii) *Decision Tree Method.* Decision tree method is to establish a flowchart of a tree structure by categorizing data variables [26]. The advantage is that the whole process of decision-making can be clearly seen, and the description is simple and easy to understand, but the disadvantage is that it is difficult to find the laws and information behind the data from the combination of multiple variables.
- (iv) *Rough Set Method.* Rough set method is specially used to deal with vague and inaccurate data analysis and statistics. It was proposed by a Polish mathematician in the 1980s. It can be used to deal with issues such as data correlation discovery and data meaning evaluation. The advantage is that the algorithm is simple, but the disadvantage is that it cannot directly deal with the problem of continuity, and the discretization of attributes is required first.
- (v) *Fuzzy Set Method.* Fuzzy set method, as the name suggests, is to perform fuzzy analysis, fuzzy prediction, and fuzzy evaluation on problems and data. The more complex the general problem, the higher the ambiguity [27, 28].
- (vi) *Association Rules Law.* The association rule method is to first find out the association between the problems and then mine the association rules that meet the minimum support and minimum credibility from the database.

2.2. *Book Copyright Information Management and Decision-Making.* With the continuous development of information technology and its implementation and application in various

aspects, traditional libraries are increasingly moving towards digital development and management models. The combination with Internet technology allows the public to better use the rich resources of the library to obtain as much knowledge and information as possible through the Internet, but at the same time, the Internet has the characteristics of fast information dissemination, wide dissemination, and large dissemination volume. Often some books and works are involved in copyright disputes. There are a large number of potentially valuable but unpublished information resources, various academic treatises, electronic publications, news materials, and syllabus on the Internet. When the library compiles these information resources into electronic information for readers to learn, it should pay attention to the protection of its copyright and intellectual property rights and should clearly indicate to readers and the public which materials can be used only after obtaining permission and which cannot be cited. In short, in the process of digital construction and management, libraries should respect the copyrights of others in accordance with the law and avoid infringement [29].

2.3. Application of Data Mining Technology in Book Copyright Information Management and Decision-Making. With the progress of the times and the development of science and technology, digital libraries have gradually become the mainstream development direction of libraries. As the main storage place for books, the library has a huge amount of book materials and information data, and with the continuous increase of social demand, the library's database is also expanding, which makes the library's book management work complicated and arduous. The consumption of human resources has also become huge. The management and decision-making of book copyright information is essentially a process of book management, and the effective management of book work can be applied to data mining technology. The information in the library book information management system is very complex, manual processing is very slow, and the efficiency is low. Using data mining technology to analyze and process book information can well reduce the complexity and low efficiency of data processing. When data mining technology is applied to book management work, SQLServer2005 is usually used as a tool. The specific implementation includes the following steps:

(1) Choose data mining technology tools

When using data mining technology to perform data mining on objects, we must first choose the auxiliary tools of data mining technology. Usually, we use SQLServer2005 as a tool. SQLServer2005 has good scalability and the ability to independently process data and can independently simplify data processing. It can run on multiple systems such as Windows at the same time and has good compatibility. In the face of highly complex data, its good performance makes it easy to solve.

(2) Collect and clean up data

After data mining technology is selected, it is necessary to collect and filter data. First, enter the book

information database, summarize all the data into a table, and then use the SQLServer2005 tool to simplify the selected data, and finally draw a simple table, and process the table to make the table manageable.

(3) Mining book copyright information data

After preliminary processing and summarization of the data, the form is embedded back into the book information management system, through the automatic classification and summary of the books, to find and complete the mining of relevant data information about the management and decision-making of book copyright information.

(4) Analyze and summarize the data obtained

After completing data mining, sort out and analyze the data information obtained, and summarize the hidden information and laws behind the data. First, compare the data obtained, and analyze the characteristics of book copyright information management and decision-making through systematic comparison, what are the types of book copyright, which books can be cited and which can only be cited after obtaining the author's permission. After finding out these regular pieces of information, it is convenient for the library to check the images on the Internet.

3. Application of the Rule Association Method of Data Mining Technology in the Management and Decision-Making of Book Copyright Information

As mentioned above, data mining technology includes neural network algorithm, genetic algorithm, decision tree method, fuzzy set method, rough set method, association rule method, etc., in order to analyze specifically and in-depth how the data mining technology can realize the copyright information of books. For data mining of management and decision-making, this article uses the association rule method to conduct application experiments on book information management and compares the Apriori algorithm in the association rule method with the improved Apriori algorithm to discuss the best way of data mining for book information management.

3.1. Rule Association Law. The rule association method is a method to find the relationship between different data that meet the given support and credibility threshold in a given database, and it is one of the common methods for data mining. Usually, the data set for association rule mining is represented by M , $M = \{t_1, t_2, \dots, t_n\}$, $t_k = \{i_1, i_2, \dots, i_p\}$ ($k = 1, 2, \dots, n$) represents a transaction, and the element i_j ($j = 1, 2, \dots, p$) in t_k represents a project. Suppose the expression of the association rule is $X \Rightarrow Y$, where $Y \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. Then, the ratio of the transaction set consisting of X and Y transactions to the transaction set consisting of all transactions is the support degree

of the association rule $X \Rightarrow Y$ in the transaction database M , denoted by $S(X \Rightarrow Y)$; that is,

$$S(X \Rightarrow Y) = \frac{|\{T: X \cup Y \subseteq T, T \in M\}|}{|M|}. \quad (1)$$

The ratio of the transaction set consisting of the number of transactions of X and Y to the transaction set consisting of the number of transactions of X is called the credibility of the association rule $X \Rightarrow Y$ in the transaction set, denoted by $C(X \Rightarrow Y)$, namely,

$$C(X \Rightarrow Y) = \frac{|\{T: X \cup Y \subseteq T, T \in M\}|}{|\{T: X \subseteq T, T \in M\}|}. \quad (2)$$

Given a transaction set M , the task of data mining association rules is to mine association rules whose credibility and support are greater than the minimum credibility and the minimum support, and the parameters describing association rules generally have four, namely, confidence, support, expected confidence, and action. Support and trust of association rules are two important concepts to describe the degree of action. Support is an association rule that is used to measure whether the whole data set is statistically significant, to measure the credibility of association rules, and to help support the degree of all the items of this association rule. Users are generally interested only in association rules with high levels of support and confidence.

3.2. Apriori Algorithm and Its Algorithm Flow. Apriori algorithm is the most classic algorithm in association rules, and it is also the most influential method in mining data association rules. It uses an iterative method to search layer by layer, which increases the frequency of frequent itemset layers. When compressing the search space, it uses a character called Apriori. Generally, the idea of the Apriori algorithm is to first form the candidate item set into a specific size, then scan and calculate to form a database, and then determine whether the candidate item set is frequently used according to the database information. The specific process is as follows:

- (1) Scan the data of all databases, calculate the number of occurrences of each item, count the first candidate item set C_1 , and then give the predetermined minimum support L_1 of the first candidate item set.
- (2) Calculate the second candidate item set C_2 according to $L_1 * L_1$, and then calculate the number of occurrences of each element in C_2 according to the predetermined minimum support L_2 of the second candidate item set.
- (3) Repeat the above steps until L_K is produced.
- (4) Delete the subset of C_j ($j = 3, 4, \dots, k$) in the candidate set C_1, C_2 . The realization of this part has two steps of connection and pruning:
 - (1) The connection step: $C_1 = I$ represents the item in the transaction, the first item set L_1 is obtained by scanning the database data, and C_2 is obtained from $L_1 * L_1$, and then the database is scanned to obtain

L_2 , and then C_3 is obtained from $L_2 * L_2$, and so on. When the loop reaches the k th time, C_k is obtained from $L_{k-1} * L_{k-1}$. If $C_k \neq \emptyset$, then continue to scan the database to obtain L_k ; otherwise, the algorithm ends.

(2) The pruning step: obtain C_k at the k th scan, and obtain L_{k-1} from the $(k-1)$ scan as the item connection set, and then obtain the item set through the connection transaction, because its subset is not a frequent item set, so the $(k-1)$ item set is not a frequent item set, so it needs to be trimmed or deleted.

- (5) Use a recursive method to deduce all frequent item sets of transactions in the Apriori algorithm, as follows:

$$\begin{aligned} L_1 &= \{\text{large1 - itemsets}\}; \\ \text{for } (k = 2; L_{k-1} \neq \emptyset; k + 1); \\ C_k &= \text{Apriori - gen}(L_{k-1}). \end{aligned} \quad (3)$$

Find a new candidate set C_k , start execution for all speeds $t \in M$, $C_t = \text{subset}(C_t, t)$ belong to the candidate set C_t of transaction t , do C_t for all C_s belonging to count+, and then end.

3.3. Improved Apriori Algorithm. From the above Apriori algorithm, we can see that it has obvious shortcomings: firstly, the number of project candidate sets is too large, which will affect the calculation efficiency and reduce the effectiveness; secondly, the database needs to be scanned multiple times, which is complicated, particularly time-consuming and inefficient. Finally, the workload of support counting becomes very complicated. Therefore, this paper proposes an improved Apriori algorithm, which reduces the number of scans of the database. Based on the relational algebra theory, it only needs to scan the database once, which greatly reduces the calculation time and improves the calculation efficiency. The improved algorithm also has good parallelism and scalability. The specific algorithm is as follows:

- (1) Set the relationship matrix

Let H represent transaction database, $T = \{t_1, t_2, \dots, t_m\}$ and $I = \{i_1, i_2, \dots, i_3, \dots, i_n\}$ represent transaction set and item set, matrix

$$R = [r_{kj}] = \begin{cases} 1, \\ 0. \end{cases} \quad (4)$$

Among them, $k = 1, 2, \dots, m$; $j = 1, 2, \dots, n$ represents the binary relationship matrix from T to I :

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix}. \quad (5)$$

Among them, the value r_{kj} of the j th item included in the k th transaction is 1, and the value r_{kj} of the j th

item not included in the k th transaction is 0. Thus, the support degree of the j th attribute of the first item set is $\sum_{k=1}^m r_{kj}/m$.

(2) Improve algorithm content

First, generate the first frequent itemset, input data: minimum support threshold t and relation matrix R ; output data: the first frequent itemset $\{j\}$.

For $j = 1, 2, \dots, n$

$f_j = 0$

For $i = 1, 2, \dots, m$

$f_j = f_j + r_{ij}$

$f_j = f_j/m$

For $j = 1, 2, \dots, n$

If ($f_j > t$)

Then output the first frequent item set $\{j\}$

Assuming that the first frequent itemset set is D , where d_i stores the first frequent itemset corresponding to it, $i = 1, 2, \dots, s$; that is, there are s elements in $\{d_i, d_j\}$, then the support for the second frequent itemset of is $\sum_{p=1}^m (r_{pa_i} \text{ and } r_{pa_j})/m$.

3.4. Specific Application of the Association Rule Method in Book Copyright Information Management and Decision System. In the face of such a huge library, how to quickly and accurately find the book they want from the huge amount of books has always been plagued by this problem. The library of colleges and universities has a huge collection of books, its library management system is very extensive, and it stores a lot of information. In order to realize the effective management of book information, improve the efficiency of book management, and help teachers and students to retrieve and find the books and materials they need faster and better, many colleges and universities have introduced digital mining technology into the library book information management system. Among them, the management of books is not only convenient for people to find, the copyright protection of books is also part of the library management work. In order to explore how data mining technology realizes the effective management of library book information, this article takes a university library in our province as an example to investigate how it applies digital mining technology to library book information management and explores its application effect situation. The specific application situation and effect will be discussed and introduced in detail in the fourth section of this article.

4. Application of Data Mining Technology in Book Copyright Information Management and Decision-Making System and Its Effect Analysis

In the third section of this article, we specifically introduced the association rule method of data mining technology and its Apriori algorithm and improved Apriori algorithm and cited a specific case. Taking a university library in our province as a specific example, we explored the data mining

technology in its application in library book information management. In this section, we will introduce the application of Apriori algorithm and improved Apriori algorithm in the book information management of the university library in detail and carry out specific analysis and discussion on its application effects.

4.1. Data Mining of Book Information and Establishment of Data Warehouse. After investigation, the library of the school has a history of more than 60 years. The library has a rich collection of books and a large and extensive library management system. Therefore, we should first sort and summarize the original book information database of the school library, establish a new data warehouse, and then use data mining technology to conduct data mining and cluster analysis on the data information in the data warehouse. Data warehouse is a kind of data storage and data organization. It is responsible for providing data sources. It is to further extract, process, and integrate data in the original database, and it is constantly updated as the content of new data increases. The structure of the data mining system is shown in Figure 1.

Table 1 presents the information of some books borrowed by the students of the School of Economics and Management of the university.

Use the association rule method to perform data mining on the data in Table 1. The calculation shows that $S(1, 2) = 0.14$, $S(1, 3) = 0.25$, $S(1, 4) = 0.14$, $S(2, 3) = 0.36$, $S(2, 4) = 0.38$, and $S(3, 4) = 0.79$; analysis shows that $S(3, 4)$ has the highest degree of support, which means that the probability of both "Management Practice" and "Management: Tasks, Responsibilities, Practices" students is 79%, because these two courses are professional books for the students of the School of Economics and Management. Although "Economic Management" has the largest number of borrowers, the possibility of borrowing the other three books at the same time is relatively low. After further analysis of the data, it is found that most of the borrowers of the book are lower-grade students.

4.2. Comparison of the Effect of Apriori Algorithm and Improved Apriori Algorithm. Taking the book borrowing data warehouse of the university library as the original database, Apriori algorithm and improved Apriori algorithm were used to conduct association rule data mining and analysis, respectively. Data mining was conducted on the 25697 book borrowing detailed catalogs stored in the data warehouse in the past two years, from which we obtained 7,905 book borrowing detailed catalogs, which are related to the top 50 books that have been borrowed. The comparison of the effects of the Apriori algorithm and the improved Apriori algorithm is shown in Table 2 and Figure 2.

It can be seen from Table 2 that compared with the original Apriori algorithm, the improved Apriori algorithm has significantly improved the efficiency of data mining. As the number of transactions in the data warehouse increases, the superiority of improving the efficiency of the Apriori algorithm in the calculation of frequent itemsets will be

FIGURE 1: Data mining system structure.

TABLE 1: Four books on economics and management.

Number	Book name	Total number of loans	Number of students borrowed from the School of Economics and Management
1	Principles of Economics	251	180
2	Management	163	123
3	Management Practice	120	95
4	Management: Tasks, Responsibilities, Practices	96	80

TABLE 2: Comparison of the two algorithms.

Number of transactions	Association rules	Time (s)	
		Apriori	Improved Apriori
1200	26	36	41
1567	48	78	83
1890	47	167	156
2145	48	268	236

more and more manifested. Through experimental comparison, the minimum support degree is 10 and the minimum credibility is 0.4. It can be seen from Figure 2 that when the transaction set is below 1400, the original Apriori algorithm is better, but as the number of transaction set records exceeds 1400, the performance effect of the improved Apriori algorithm is significantly higher than the original Apriori algorithm.

4.3. Implementation of the Improved Apriori Algorithm in Book Information Management. In order to further explore and improve the outstanding performance of Apriori algorithm in book information management, we once again use the university's book borrowing database as a sample for data mining. The types and times of book borrowing in each

college are shown in Figure 3. The borrowing ratio is shown in Figure 4.

From Figures 3 and 4, we can see that literature, history, and geography books have a relatively high number of borrowing in each college, and they account for a relatively large proportion of the total number of books borrowed, of which literature books account for the largest, reaching 29.2%. Due to the focus of majors, the corresponding books borrowed by the colleges are more biased towards the majors they have studied. Among them, the School of Economics, the School of Foreign Languages, the School of Art, the School of Agriculture, and the School of Computer Science are the most prominent. In Figure 4, philosophy, politics and law, military, natural science, and agriculture books are the least borrowed, accounting for only 0.9%, 1.25%, 0.5%, 1.5%, and 1.2% of the total book borrowing volume, respectively.

FIGURE 2: Performance comparison of the two algorithms.

FIGURE 3: Book lending situation of each college.

Military books have the least amount of borrowing. Through the analysis of the data, it shows that the reading hobby of the students in this school tends to be historical and literature and that literature books have a faster turnover and

larger circulation than science books; it also reflects that the number of natural science books in the school library is less than that of history. Literary books, and quite a few are outdated, should be supplemented in time.

FIGURE 4: Proportion of various books borrowed.

5. Conclusions

With the development of social economic and the improvement of people's living standards, more and more people are beginning to pay attention to the improvement of spiritual realm and the cultivation of knowledge and sentiment. Books have been called people's spiritual food since ancient times, and they are used in people's daily leisure and entertainment life; they occupy a very heavy weight. As a place for teaching and educating people, the library has the largest and most complex collection of books. How to realize the effective management of book information in the library has always been one of the most troublesome problems for universities.

The rapid development of information technology and network technology has brought tremendous convenience to people's production and life. With the continuous improvement of network and informatization, the effective management of library information has become possible. Digital mining technology is an emerging big data processing and analysis method, which has the ability to quickly process and analyze data, and can find the information and laws hidden behind the data. Applying data mining technology to the book information relation system can realize the effective management of book information.

Among the various algorithms of data mining technology, the association rule method shows great advantages, but its original Apriori algorithm has certain limitations. It

only shows great advantages in a small-scale transaction set project, while the number of the library information materials is generally huge and complicated. The original Apriori algorithm cannot meet the ideal requirements for the data processing of its database. For this reason, this research proposes an improved Apriori algorithm. Improved Apriori algorithm shows better performance advantage in the face of big data calculation.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the Major Project of Beijing Social Science Foundation "Research on Financial Support System Adapting to the Coordinated Development of Strategic Emerging Industries in Beijing-Tianjin-Hebei" (no. 20ZDA11).

References

- [1] R. Kanniga Devi, M. Gurusamy, and P. Vijayakumar, "An efficient cloud data center allocation to the source of requests," *Journal of Organizational and End User Computing*, vol. 32, no. 3, 2020.
- [2] R. Ye, "Application of web data mining in information management," *Science and Technology Plaza*, vol. 000, no. 003, pp. 146–149, 2017.
- [3] X. Tian, "Analysis of intellectual property protection in library service and management," *Journal of Guizhou Commercial College*, vol. 031, no. 003, pp. 69–72, 2018.
- [4] A. Abu Sirhan, K. M. Abdrabbo, S. Ahmed Ali Al Tawalbeh, M. Hamdi Ahmed, and M. Ali Helalat, "Digital rights management (DRM) in libraries of public universities in Jordan," *Library Management*, vol. 40, no. 8/9, pp. 496–502, 2019.
- [5] S. L. Guo and L. N. Han, "Application and exploration of big data mining in clinical medicine," *Chinese Medical Journal*, vol. 129, no. 006, pp. 731–738, 2016, https://pubmed.ncbi.nlm.nih.gov/?term=Li+TL&cauthor_id=26960378https://pubmed.ncbi.nlm.nih.gov/?term=Zhang+Y&cauthor_id=26960378.
- [6] J. Wu, S. Guo, J. Li, and D. Zeng, "Big data meet green challenges: big data toward green applications," *IEEE Systems Journal*, vol. 10, no. 3, pp. 888–900, 2016.
- [7] S. Ren, F. Shen, H. Xu, M. Li, and J. Wu, "The application of data warehouse and data mining in fracturing engineering system," *IOP Conference Series: Earth and Environmental Science*, vol. 170, no. 2, Article ID 022080, 2018.
- [8] H. S. Behera and D. P. Mohapatra, *Harmony Search Algorithm Based Optimal Tuning of Interline Power Flow Controller for Congestion Management*, Springer, Berlin, Germany, Chapter 16, pp. 173–182, 2016.
- [9] S. E. Apitz and S. Agius, "Anatomy of a decision II: potential effects of changes to Tier I chemical approaches in Canadian Disposal at Sea program sediment assessment protocols," *Integrated Environmental Assessment and Management*, vol. 13, no. 6, pp. 1072–1085, 2017.

- [10] C. F. D. L. Hoz, E. Ramos, A. Puente et al., "Ecological typologies of large areas. An application in the Mediterranean Sea," *Journal of Environmental Management*, vol. 205, pp. 59–72, 2017.
- [11] L. Li and J. Zhang, "Research and analysis of an enterprise E-commerce marketing system under the big data environment," *Journal of Organizational and End User Computing*, vol. 33, no. 6, 2021.
- [12] N. Japkowicz and J. Stefanowski, *An Overview of Concept Drift Applications*, Springer, Berlin, Germany, Chapter 4, pp. 91–114, 2016.
- [13] S. Kudyba and K. Lawrence, "Enhancing information management through data mining analytics to increase product sales in an e-commerce platform," *International Journal of Electronic Marketing and Retailing*, vol. 2, no. 2, pp. 97–104, 2017.
- [14] H. Li, "Application of data mining technology in hospital management information system," *Digital Technology and Application*, vol. 000, no. 009, pp. 228–229, 2017.
- [15] Q. Lu, "The application of GIS digital mapping technology in mining geological survey," *Engineering Construction and Design*, vol. 000, no. 006, pp. 257–258, 2018.
- [16] G. Zhang, "Application of Multimedia data mining technology in the development and utilization of digital library collection resource," *Journal of Agricultural Library and Information Science*, vol. 030, no. 003, pp. 61–64, 2018.
- [17] J. Kamps, G. Tsakonas, Y. Manolopoulos, Y. Iliadis, and L. Karydis, "Research and advanced technology for digital libraries," *Mining Semantic Patterns for Sentiment Analysis of Product Reviews*, vol. 10450, Springer, Lecture Notes in Computer Science Chapter 30, pp. 382–393, 2017.
- [18] L. Mao, Y. Zhu, Y. Wang, Y. Liu, L. Li, and F.-p. Chiang, "An improved digital volumetric speckle photography technique with X-ray microtomography and its applications to investigating strain localization in red sandstone," *Rock Mechanics and Rock Engineering*, vol. 53, no. 3, pp. 1457–1466, 2020.
- [19] J. Wang, L. Bi, L. Wang, M. Jia, and D. Mao, "A mining technology collaboration platform theory and its product development and application to support China's digital mine construction," *Applied Sciences*, vol. 9, no. 24, p. 5373, 2019.
- [20] Y. Gao, "Discussion on safe and efficient mining technology driven by digital information," *World Nonferrous Metals*, vol. 000, no. 008, p. 81, 2017.
- [21] W. . Qiu, "Application of digital technology in product design based on data mining," *IPPTA: Quarterly Journal of Indian Pulp and Paper Technical Association*, vol. 30, no. 7, pp. 517–523, 2018.
- [22] N. Fuhr, L. . Kovács, T. Risse, and W. Nejdl, "Research and advanced technology for digital libraries," *Person-Centric Mining of Historical Newspaper Collections*, vol. 9819, Springer, Lecture Notes in Computer Science Chapter 25, pp. 320–331, 2016.
- [23] O. Evsutin and Y. Meshcheryakov, "The use of the blockchain technology and digital watermarking to provide data authenticity on a mining enterprise," *Sensors*, vol. 20, no. 12, p. 3443, 2020.
- [24] G. Tan, "Research on the application of mining technology in software engineering," *Digital Technology and Application*, vol. 037, no. 004, pp. 114–122, 2019.
- [25] F. Song, Q. Zhou, and M. Tang, "Application of digital integration technology in mine safety production," *Modern Mining*, vol. 000, no. 006, pp. 182–184, 2019.
- [26] B. I. Schneider, B. R. Miller, B. V. Saunders, and V. Bonita, "NIST's digital library of mathematical functions," *Physics Today*, vol. 71, no. 2, pp. 48–53, 2018.
- [27] X. Wang, H. Schneider, and K. R. Walsh, "A predictive analytics approach to building a decision support system for improving graduation rates at a four-year college," *Journal of Organizational and End User Computing*, vol. 32, no. 4, 2020.
- [28] J. Wu, K. M. Williams, H. H. Chen et al., "CiteSeerX: AI in a digital library search engine," *AI Magazine*, vol. 4, no. 3, pp. 2930–2937, 2017.
- [29] M. N. Masrek and J. E. Gaskin, "Assessing users satisfaction with web digital library: the case of Universiti Teknologi MARA," *The International Journal of Information and Learning Technology*, vol. 33, no. 1, pp. 36–56, 2016.