Application of Data Mining Technology in Book Information Management

Kun Dong\textsuperscript{1,2,*}

\textsuperscript{1}School of Information Management, Wuhan University, China, 430074
\textsuperscript{2}South Central University for Nationalities Library, China, 430074

*Corresponding author e-mail: dongkun@whu.edu.cn

Abstract. With the development of big data mining technology, it plays a very important role in the university library management system. It effectively solves the problem of data searching in traditional library management. This article mainly analyzes and studies the application of data mining technology in libraries.

Keywords: Data Mining, Association Rules, University Libraries

1. Introduction

Nowadays, libraries need to provide accurate, timely and personalized services to teachers, students and readers. University libraries must use computers to improve service quality, and at the same time we must obtain truly effective and reliable data in the ever-expanding and increasingly complex system. Using a convenient and efficient way to manage online books is an urgent problem for the library management department.

2. Data mining technology

2.1. Implementation plan of data mining

Driven by the rapid development of the Internet of Things and computer technology, the application of big data has been extended to many industries. It not only subverts people's cognition of data in the traditional sense, but also induces data acquisition, storage, analysis, mining, and visualization technologies. change. At the same time, the current human production and lifestyle will also undergo tremendous changes due to big data and related technologies. "Huge amount of data" is only the superficial feature of big data. Its brand-new data processing mode and its decision-making power and insight that cannot be accomplished by traditional tools in a short time are the more core meaning of big data [1]. However, the original data often contains impurities and a lot of interference information, and most of the time these data are vague and have no obvious rules. Only by applying certain technical
means to filter out the impurities and interference information in the existing data information can you obtain truly valuable data, so that more scientific decisions can be made based on big data.

The main methods of data mining include association rule analysis, cluster analysis, classification analysis, time series pattern prediction, and deviation analysis. The following is an introduction to the association rule analysis and commonly used Apriori algorithm involved in this article [2].

2.2. Association rule analysis

In the era of big data, a large amount of seemingly chaotic, unruly data often has deep potential connections behind it. The technology of finding associations or dependencies between data from large amounts of data is called Association Analysis (Association Analysis). The basic principle of association rule analysis is as follows.

Let $D=\{t_1, t_2, \ldots, t_k, \ldots, t_n\}$ be the data set to be mined, and let $t_k=(i_1, i_2, \ldots, i_m, i_p)$ be a transaction, where $\text{im}(m=1, 2, \ldots, p)$ is an item, $X$ is defined as an item set, and its support number is defined as the number of items contained in data set D. The number of transactions is denoted as $x$, and the support degree of $X$ is denoted as support($X$).

$$\text{support}(X) = \frac{x}{|D|} \times 100\% \quad (1)$$

Where $|D|$ is the number of transactions in data set D, and the minimum support threshold is defined as min sup. When support($X$) $\geq$ min sup, $X$ is called frequent itemsets, and vice versa, $X$ is called infrequent itemsets [3].

Note that $X$ and $Y$ are itemsets in data set D. If they exist, support($X$) $\geq$ support($Y$). If $X$ is an infrequent item, then $Y$ is also an infrequent item, and if $Y$ is a frequent item, then $X$ is also For frequent items. If the item set $X \cap Y = \emptyset$, then $X \Rightarrow Y$ is called an association rule, $X$ is the premise of the association rule, and $Y$ is the conclusion of the association rule. Its support is the support of XUY, denoted as support($X \Rightarrow Y$). Let the confidence of the association rule $X \Rightarrow Y$ be confidence($X \Rightarrow Y$), then the calculation of the confidence of the association rule $X \Rightarrow Y$ is shown in equation (2).

$$\text{confidence}(X \Rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} \times 100\% \quad (2)$$

Define the minimum confidence threshold as min conf. If the association rule of $X \Rightarrow Y$ satisfies support($X \Rightarrow Y$) $\geq$ min sup and confidence($X \Rightarrow Y$) $\geq$ min conf, the association rule $X \Rightarrow Y$ is considered to be a strong rule. Otherwise, it is a weak rule [4].

In the process of data mining, the goal is to find all the strong association rules hidden in the data set D, that is, to find the frequent itemsets of the corresponding itemsets of the association rules $X=Y$.

It can be seen from Figure 1 that data mining based on association rule analysis is the process of mining the minimum support and minimum credibility set by the user through the interaction of two algorithms.
3. Data mining model based on association rule analysis

Librarians and readers are the two major users of the library management system in colleges and universities. Only when administrators establish enough contact and interaction with readers can they better understand the needs of readers for books [5]. After understanding the needs of readers, can they further optimize books. The types and quantity of books in the library reduce the waste of resources and improve the service function of the library. Once a university library is established, the collection will inevitably continue to increase, and the scale of library management will also increase accordingly. While the data of library management is gradually increasing, it will also show the following characteristics.

(1) The complexity of data information: A large amount of data will be generated in the process of reader retrieval, reading, and management of books by administrators. At the same time, students, as the main users of university libraries, have very limited time in school. The library will have new users joining and old users exiting, which means that the data of book management will continue to increase.

(2) The obscurity of the information relationship: The accumulation of massive data will obscure the real regularity behind the existing data. Only with the help of effective algorithms to analyze and process the huge volume of data can we discover its internal content. More valuable information.

(3) The relevance of information: There will be a certain relevance between the user's own information and the retrieval information of the book, but this relevance is relatively simple. On the basis of understanding the potential of massive data, explore Only by extracting the truly valuable information behind the data can we further find the correlation between the data, and with the support of sufficient depth correlation, we can better complete the book management work [6].

The complexity of library management data information The obscure nature of information relationships and the relevance of information put forward higher requirements for the connection between librarians and users, and traditional book management methods cannot establish such a strong, efficient and accurate This provides a stage for the application of data mining technology in the library management system.

4. The demand for data mining in library management
The data mining in the library management of colleges and universities studied in this article mainly has three levels: "data preprocessing", "associated data mining" and "model visualization analysis". In the process of data mining, we mainly focus on "analysis of readers' borrowing behavior patterns". Personalized service work for administrators" and "document shelf management work"

4.1. Hierarchical structure of data mining

The main content of the hierarchical structure of data mining in university library management is shown in Figure 2 [7].

![Hierarchical structure diagram of data mining in university library management](image)

(1) The data preprocessing stage is the data collection stage. First, the generated data is converted and integrated from the database of the library management system, and then the integrated data is exported into a format that can be used by the data mining system. The data is saved in the library management data mining library

(2) In the association rule mining stage, based on association rules and appropriate algorithms (such as Apriori algorithm), the library management data prepared in the first stage is excavated and processed.

(3) In the visual analysis stage of the model, first determine the data mining goals as book borrowing behavior, personalized service and document shelving, then select the book management mining task parameters to mine the book management data, and finally output the visual pattern rules [8].

4.2. The needs of data mining

(1) Analysis of borrowing behavior patterns: First, analyze the reader group according to the reader's retrieval record borrowing habits and requirements, etc., to obtain the popularity of the book among the readers; secondly, combine the reader's identity information and the retrieval and borrowing information
The characteristics of the reader group are analyzed, and the reader's search and borrowing behavior rules are obtained with the help of clustering data.

(2) Personalized service work: First, analyze the relationship between the reader's identity information and its retrieval and borrowing records, study the specific needs of different readers, and provide readers with personalized service creation data support; then analyze the different readers' The borrowing time and borrowing sequence of different types of books, analyze readers' reading habits, and provide data support for the library to rationally arrange the book layout and provide readers with appropriate books in different time periods.

(3) Library document shelf analysis: Based on the analysis of borrowing behavior patterns and personalized service work, data mining technology is used to analyze the popularity of different books and the borrowing records of different books to provide library documents. The reasonable arrangement of shelves and the improvement of book utilization rate provide data support.

5. Application of Data Mining Technology in Book Management

This article takes a university library as an example. The university was established in 1952 and currently has more than 230,000 students. The library has a total area of 41,765 square meters. As of the end of 2017, the total number of books in the library is 3,699,800, of which 243.26 million are paper books. 126.72 million e-books, 122 digital resources in Chinese and foreign languages Part of the reader's borrowing information is selected from the university's library management system as an example, as shown in Table 1, where "Y" means the book is lent, and "N" means the book is not lent [9].

**Table 1.** Book borrowing records

From the book borrowing records shown in Table 1, we can draw the following conclusions
(1) When readers borrow in the library, there is also a strong correlation between different books. For example, readers who borrow "Python Basic Course" will also compare "Introduction to Data Mining" and "Data Mining Concepts and Techniques" Interested, and readers of "Linear Programming" will often borrow "Combinatorial Mathematics" and "MATLAB Collection";

(2) Different readers have different needs for different books. Since the collection of any book in the library is limited, the mining of frequent books is crucial to the optimization of the number of book collections in the library [10].

6. Conclusion

With the advent of the information age, Internet technology has been widely applied to people's lives. The library information management system has become the main medium for the library to provide services to readers. This has an important impact on the development of the library.

Acknowledgments

Teaching and Research Project of South-Central University for Nationalities in 2020 (Project No. JYX20123), "Research on the Construction of Document Resources Guarantee System Oriented to Teaching and Research under the Background of Double First-Class".

References

[1] WU Jing. Research on Application of Data Mining in Library Management [J]. Journal of Agricultural Library and Information, 2015,27 (12):54-57.

[2] LI Huiyan. Application of Data Mining Technology in Library Management of Higher Vocational Colleges [J]. Journal of Agricultural Library and Information, 2015,27(3):65-68.

[3] Zhang Qing, Li Qi, Cheng Pengzhou. Research and implementation of mobile book management system in the era of big data [J]. value engineering, 2018,37(26):44-46.

[4] Zhang Xianheng. Application of Data Mining Technology in Library Management [J]. Science and Technology Information, 2012:260+262.

[5] CHEN Jun. Development of Library Information Management System Based on Data Mining Technology [J]. Electronic Technology and Software Engineering, 2018:184.

[6] XUE Jian. Research on Information Management of College Books Based on Data Mining Technology [J]. Information Technology, 2019:83-86+91.

[7] Pan Fang, Huang Fengkai. Development of Library Information Management System Based on Data Mining Technology [J]. Automation & Instrument, 2019:184-187.

[8] Dong Yan. Application Analysis of Data Mining Technology in Information Management of University Library [J]. Science, Technology and Economy Guide, 2017:23.

[9] Zhang Yongqiang. Application of Data Mining Technology in University Library Management [J]. Journal of Huaibei Vocational and Technical College, 2017.
[10] Ge Xiaobin, Liu Jie. Application of Data Mining Technology in University Library Management [J]. Journal of Hunan City University: Natural Science Edition, 2016: 50.