Research and Application of Data Mining Technology in Library Office Information Construction

Weifeng Wang¹, Luyang Meng¹, Lei Wu¹, Jing Zhang¹

¹Library, University of Jinan, Jinan, Shandong Province, 250022, China

Abstract. With the rapid development of the Internet, the amount of data related to users has shown a rapid increase. At the same time, in the library office system, the problem of how to effectively use and process massive data needs to be solved urgently. Data mining technology, as a technology combining machine learning, pattern recognition, and data visualization, can effectively solve many problems in the process of library office information construction. The data mining technology mainly includes many analysis methods such as association analysis, data segmentation, and deviation analysis. The mining results of these algorithm models are presented to users in a visual way, and the results of the rules, trends and concepts discovered can be applied to During the application of the system.

1. Introduction

With the continuous development of the era of big data, data mining technology is widely used in various systems. In fact, data mining technology is not a separate discipline and technology. It has integrated knowledge in many fields such as computer, mathematics and information. Data mining, as the name implies, is to find the most useful information for this matter in a large number of databases. Of course, because useful information is hidden in a lot of useless, noisy, and even incomplete data, it is necessary to "dig" this technology to find useful potential information. Before the data mining, we did not know the specific steps of the entire mining process, the problems that may be encountered, and the quality of the mining results. The final result in the data mining process is not just numerical, it may be laws, models, concepts, etc. The current data mining technology used in application systems usually incorporates multiple aspects of machine learning, pattern recognition, neural networks, data visualization and other technologies, making its experiments and application results good.

Whether it is business, medical or education, it is actively using data mining technology, not only because data mining technology can solve problems that cannot be solved in traditional mathematical knowledge, but also because of the huge economy hidden behind data mining technology interest. At this stage, the most widely used field of data mining technology is the financial field. The prediction of risks and the prediction of trends cannot be separated from the participation of data mining technology. With the increasing amount of data, more and more practical problems frequently appear. At this time, traditional mathematical models appear to have many disadvantages. And the model algorithm in data mining technology is more and more prominent in its many advantages, which makes massive data and its ability to analyze and process massive data become extremely important. First, data mining tends to be diversified, comprehensive, and integrated in terms of application fields, methods of use, and types of results. And all aspects of data mining can continue to develop in depth. At the same time, for a large amount of private information in the database, data mining will also process and apply the information under complete confidentiality. Because of these advantages, data mining technology has been well received in many analytical methods.
Library office information construction is the product of the continuous development of multimedia technology. With the continuous improvement of information, the office efficiency of system software has been greatly improved. Especially in the face of a huge amount of data in the library office management system, information construction has become extremely important. First of all, a reasonable and effective office information construction can greatly promote the automated processing of library affairs, save office time, and improve processing efficiency. Secondly, the informatization of book information resources will greatly increase the degree of openness of information and facilitate the exchange and sharing between libraries. Finally, the higher the digitization of the library's office system, the easier it will be to become a resource concentration center in the entire education process, which will not only facilitate students' study and reference, but also benefit the overall further development of the library. In the process of library office information construction, data mining technology must be used. On the one hand, data mining technology can find the advantages of the existing construction of the library from a large number of applied data. On the other hand, data mining technology can also obtain many disadvantages from a large amount of data, which is conducive to the continuous improvement of the next step of library construction. Therefore, the integration between data mining technology and library office information construction is essential.

2. Data mining technology
Data mining techniques include association analysis, sequence analysis, classification, prediction, cluster analysis, and time series analysis.

- **Association analysis**: Association analysis technology is mainly proposed on the basis of mathematical probability calculation formulas, which is also intended to calculate the probability of the occurrence of another event under the conditions of one event. Analysis of the correlation between things can effectively find the potential correlation function between things. The obtained correlation curve is conducive to the expansion and application of related affairs, and has high practical significance.

- **Sequence analysis**: It is not only found that the time of correlation occurrence in the same time period is valuable, but also the events that occur continuously in different time periods have high research value. If the same thing or multiple different things occur one after another after a certain time interval, it can be shown that there is a greater correlation between this thing or these things and continuous time. This analysis method, which is closely related to the continuity of time, is called a sequence analysis method. That means that this event can form a certain continuity in time, so it has great research significance.

- **Classification analysis**: Classification analysis is mainly used in the classification and grouping process of each sample. Through this kind of analysis method, the sample to be classified can be easily classified into its category according to its characteristics. At this point, there are multiple approximate categories, and only a single sample needs to be classified. There are many analysis methods for reasonable classification of samples, including: decision tree algorithm, neural network algorithm, Bayesian algorithm, etc. The experimental results of these methods vary according to the specific application environment, so they should be selected reasonably in the actual application process.

- **Cluster analysis**: The effect of the cluster analysis method is roughly similar to the principle of the classification analysis method. The final result is presented in the form of the clustered samples. Clustering first needs to effectively aggregate multiple unclassified samples according to their most significant characteristics, that is, samples of the same class have roughly the same sample characteristics, and the higher the dissimilarity between the more distant classes. However, at this time, most samples have no categories, and the work performed by clustering is to classify most samples.

- **Prediction**: In the prediction process, the characteristics of the sample are used like classification analysis and cluster analysis, but the calculation result of the algorithm is different. The result of the prediction algorithm is to predict the continuous values of the samples, and the results obtained need to be judged for accuracy. However, the result of the classification analysis method is only a category value, which is a certain value. The commonly used method in the prediction algorithm is a regression analysis algorithm.
Time series: Algorithms that are very similar to prediction algorithms are information on time series. As time changes, the sample itself changes constantly. Therefore, it is necessary to make certain predictions on the changing trend of things from the perspective of big data based on time variables. The prediction algorithm determines the continuous values of the samples, while in the time series, it only makes a certain prediction for the change trend. And the changing trend of things often has a certain regularity and periodicity, so it is of great significance to effectively predict this.

The methods of data mining mainly include: prediction model method, data segmentation method, association analysis method and deviation analysis method. Among them, there are many algorithms included in the prediction model method. Neural networks, decision trees, particle swarms, support vector machines, combination classification algorithms, Bayesian classification algorithms, and classification algorithms based on association rules are all prediction model algorithms. As shown in Figure 1. Because of its clear hierarchical structure, the decision tree algorithm is suitable for further exploration and layering and is used in many systems. Moreover, the decision tree algorithm has a better processing effect on high-dimensional data, which is also a major feature. Neural networks are used to simulate the neural network structure in the human brain to connect and explore data with each other. The connections between neurons in the human brain are intricate but highly logical, so it is very appropriate to use them to describe the connection between transactions. The regression algorithm mainly includes multiple linear regression algorithm and logistic regression algorithm, the latter is more widely used in data mining. In the data mining process, classification algorithms based on association rules are often used to find frequent patterns in data sources. Frequent patterns can effectively discover the association rules between transactions.

![Main analysis methods in data mining](image)

Figure 1: Main analysis methods in data mining

Although there are many types of data mining techniques, the specific steps of each method are roughly the same. The actual operation steps of the data mining technology are shown in Table 1. Before any technology can be used in practice, it must be clear what the intended purpose of this technology is after use, in order to better perform the test operation. After fully understanding the problem of the results to be continued, a detailed understanding of the data set to be used in data mining is needed. The
format of the data, the size of the data, the type of data, etc. all need to be checked. In addition, the data with large noise and the missing data need to be further supplemented and improved before real data mining operations can be performed. The data sources for data mining operations usually come from object-oriented databases, multimedia databases, heterogeneous databases, and relational databases. In the process of data mining, the final experimental results are mainly obtained through a series of model operations. Therefore, the reasonable selection and accurate application of the model becomes very important.

Table 1: Steps and explanation of the data mining process

| Steps of the data mining process | Step explanation |
|----------------------------------|------------------|
| Understand the business          | Understanding the business means transforming the actual problem that requires data mining into a definition problem that can use an algorithmic model. In response to this problem, the desired expected result is obtained. |
| Understanding the data           | That is to fully understand the data set required for this problem. This step is the foundation of the entire data mining process. As long as you fully understand the quality and type of data, you can better apply model algorithms to it to solve problems. |
| Data processing                  | For massive data, the assistance of a series of data processing tools is needed for processing. Use various processing functions to complete the work of cleaning, transforming, supplementing, and repairing data in the entire database. |
| Modeling                         | This step is a more critical step in the entire data mining process. How to choose a model algorithm, what algorithm model to choose, and how to set specific parameters in the algorithm are all issues that need to be considered urgently when modeling. |
| Model evaluation                 | Model evaluation is an important step after the modeling is completed. In this step, you need to judge each step in the modeling and the overall performance of the model. At the same time, the actual problems and expected results in the step of "understanding the business" need to be compared to ensure that the initial problems have been effectively resolved. |
| Model deployment                 | What is obtained in the model evaluation process is not the final result of the model. The results of the model also need to be displayed to users and customers, so further deployment work is needed. This link is connected to the later part of the data result display, which is to facilitate further understanding and better use of the results of data mining algorithms. |

The entire application process of data mining technology is shown in Figure 2. First, you need to apply data mining tools to model from a large amount of data related to things and use various analytical methods to obtain the expected potential laws and values. Then the obtained values and trends need to be visually displayed by visual tools. Finally, the entire mining and analysis results need to be summarized and adjusted in practical applications. For the library office information construction, the potential information that is mined from a large amount of library resource information and many user borrowing information must not only be used to timely and accurately mine the available potential information, but also to be displayed in the form of visual charts. Show it to the staff. Only through simple and clear result analysis and viewing, users and office staff can clearly understand the source of the problem, and make corresponding adjustments and improvements to the next office construction.
3. Library office information construction
Initially, the library office informatization process was faced with three major problems: the urgent need to improve communication capabilities, the slow updating of information resources and the environment, and the low degree of professionalism of management staff. The improvement of communication capabilities requires the support of the hardware environment. Therefore, during the information construction, the hardware equipment needs to be constantly updated to obtain effective protection. The slow update of information resources and environment is a more prominent problem. Aiming at the direction of resource update, the results of the data mining process provide a lot of reference directions for it. Only by analyzing the laws and trends obtained from a large amount of data, can we provide more objective support for the direction of information sharing and updating.

In the library office information system, the hardware system, book resources, office software and system data together form a complete system. The results generated during the data mining process will provide effective assistance during the deep development of the system. As shown in Figure 3, after the system is completely constructed, it mainly involves three aspects of system operation, business reporting, and deep development. Although data mining technology has also participated in the first two links, it mainly provides effective assistance in the third link. On the one hand, the results of the data mining will be displayed to the office staff in a visual form, so that they can clearly understand the contents of book lending and personnel lending in a period of time, which is convenient for the next work direction. The book borrowing situation mainly includes the number, type and feedback of borrowed books. The personnel lending situation mainly includes the history of the names, types, quantities, and returns of the books borrowed by the personnel. This information can help the office system to complete the book renewal, recommendation, and reminder work. On the other hand, according to the book and the feedback information of the borrower, the module content in the online system can be effectively and effectively supplemented, so as to provide a better experience for the office staff and the borrower.
4. Conclusions
Data mining technology has a very important position in the research and application of many systems. The mining results can not only show the good results of the application, but also summarize the shortcomings in the system and provide guidance for subsequent improvements. Library office information construction has become an important part of the overall work of the library office. The results of data mining technology can provide effective assistance for office staff in the library office information construction and provide significant help for the continuous improvement of the system.

References
[1] Li Zhongwen. On the Application of Data Mining Technology in Library [J]. Computer Fan, 2017 (1).
[2] Zhao Qiushi. Application of data mining technology in digital library services [J]. Information Recording Materials, 2019 (6): 168-169.
[3] Jiang Yin. Research on the Application of Data Mining Technology in Library Management Information System [C], 2018.
[4] Li Qing. Application Value Analysis of Big Data Mining in Library Personalized Service [J]. Changjiang Series, 2018 (22).
[5] Wang Zhiping. A Summary of New Library Technology Research [J]. Inner Mongolia Science Technology and Economy, 2017 (5): 160-161.
[6] Zhang Yuying. Discussion on Library Informationization Construction [J]. Northern Literature (Late) (1): 238.
[7] Li Hui. Research on Strengthening Library Informationization Construction [J]. Industry and Science Forum (11): 257-258.
[8] Liu Shuling, Li Meiying. Construction and Application of Library Informationization [M]. 2012.
[9] Shan Yan. Discussion on the Construction of Professional Library Resources Informatization [C], 2015 Academic Exchange Conference of the Management Science Branch of China Aviation Society. 2015.