The Construction of Smart Library Based on Data Mining Technology

Xin Zhou1,*

1Library, Nanjing Xiaozhuang University, Nanjing, 211171, China

*Corresponding author e-mail: zhouxin0724@njxzc.edu.cn

Abstract. Data mining technology is a technology that mines and analyzes data deeply. Applying data mining technology to the construction of a smart library can quickly extract information about books and documents. This provides support for the management of document resources in smart libraries. At the same time, it not only optimizes the collection structure, but also expands the service model. It improves the quality of information services.

Keywords: Data Mining, Smart Library, Book Classification

1. Introduction

With the continuous development of database and network technology, the literature resources have increased exponentially in university libraries. Readers’ information level and information requirements continue to improve. Only relying on traditional database retrieval mechanisms and statistical methods can no longer satisfy the needs of reality. Therefore, in order to avoid falling into the situation of "rich data but lack of information", the library must use data mining technology to analyze, organize and develop in-depth information. It is necessary to dig out the inner connections from the seemingly messy information, so as to guide the work of the library. This can improve the quality of service and meet the needs of readers [1].

2. The connotation of data mining technology

2.1. Definition of data mining

Data Mining (DM), also known as data mining, is to extract from a large amount of, incomplete, noisy, fuzzy, and random data that is hidden in it that people do not know beforehand but are potential and useful information and knowledge process. It is the process of exploring and analyzing large amounts of data through automatic or semi-automatic tools, with the purpose of discovering meaningful patterns and laws.
The process of data mining is the process of knowledge discovery. The knowledge that can be discovered is as follows: one is generalized knowledge, which reflects the common nature of similar affairs; the other is characteristic knowledge, which reflects the characteristics of all aspects of things. Knowledge; the third is differential knowledge, that is, the knowledge that reflects the difference in attributes between different things; the fourth is related knowledge, that is, the knowledge that reflects the relationships between things; the fifth is predictive knowledge, that is, to predict the future based on historical and current data Data; the sixth is deviating knowledge, that is, revealing the phenomenon of things deviating from convention [2]. All of this knowledge can be found at different conceptual levels. With the improvement of the concept tree, from micro to meso to macro, to meet the needs of different users and different levels of decision-making.

2.2. Methods of data mining

Commonly used data mining tools and methods include: classification, clustering, association analysis, sequence analysis and time series, deviation detection, outlier analysis, pattern recognition, visualization, decision tree, genetic algorithm, uncertainty processing, etc.

The association in association analysis refers to a certain regularity between the values of two or more variables. The appearance of one pattern means the appearance of another pattern. It is an important type of discovery in the database. The purpose of the knowledge is to explore the interrelationships hidden in the data, including check correlation, time series correlation, and causal correlation. Usually the association rules need to find out that the support and confidence are greater than or equal to the minimum support and confidence specified by the user.

Classification is to find a set of models that can describe the typical characteristics of a data set, so as to be able to classify and identify the attribution or classification of unknown data, that is, to map unknown cases to one of a certain discrete category. Clustering refers to the process by which records in the database can be divided into a series of meaningful subsets. It is different from classification and prediction. Cluster analysis only analyzes data objects without considering the constant class labels. Before clustering, the number and type of groups to be divided are not known, nor is it known according to which data item to define the group. Divide the data into different groups, make the difference between the groups as large as possible, and the differences within the groups as small as possible, organize the observed content into a class hierarchical structure, and organize similar events together. It is a reciprocal process with classification and prediction methods, and it can handle types, values, and text data [3].

2.3. The process of data mining

Data mining generally consists of several main stages: determining the mining object, data preparation, model establishment, data mining, result analysis and mining application. Data mining can be described as an iterative process of these stages, as shown in Figure 1.
3. Innovative services and innovative technologies of libraries in the information age

3.1. Current status of library development

With the acceleration of the construction of library information resources and the improvement of reader service requirements, the role of the library is gradually changing, from the initial "Cangjing Pavilion" to the later "gatekeeper" to the current "information center". In the context of the information age, libraries already have certain characteristics of big data. There are many types of library data. According to different purposes, they can be divided into collection and editing system data statistics, circulation reading system data statistics and other data statistics.

First of all, the statistical data of the acquisition and editing system includes not only the statistics of paper books, but also the statistics of acceptance, recommendation, and collections. These data are changing at any time [4]. How to effectively use these data is a question that we need to think about. Secondly, the circulation reading system data mainly includes document management statistics, document circulation statistics, reader statistics, reader distribution statistics, etc. This is a huge body of data, and the hidden value is more worthy of our exploration. Third, other data mainly include statistics on the use of library technical services by readers and statistics on the use of online resources. Only when we truly understand the needs of readers can we better serve readers. With the increasing service requirements of users, libraries must make corresponding service strategy changes based on users' service information and other data, and it is inevitable to analyze large amounts of data and tap potential value.

3.2. Library technology innovation

With the popularization of database technology, library data is becoming more and more abundant, but the traditional information management mode of the library cannot fully and effectively mine the value of data, and even more and more libraries have entered the "rich resources, skinny value" status. At this time, applying data mining technology to the construction of the library can further realize the transformation of "data->information knowledge->value" in the library data. Combining the characteristics of big data and the status quo of resources, the integration of digital resources based on user needs is also the future development direction. The visual display of knowledge graphs such as time axis and geographic axis is used to present resources to users; relying on the integrated Internet of
Things and mobile communication network As well as the Internet for dissemination, it will eventually provide users with the reception of multiple terminals such as televisions, computers, and mobile phones to fully realize common knowledge sharing.

3.3. Library Service Innovation

Technological innovation will also further enhance service innovation. Library services will gradually be promoted from passive services demanded by readers to active data analysis services. In the era of big data, library services will also be promoted in terms of service methods, channels, and models. Will all change. The expansion of library resource data volume, the improvement of service quality, and the transformation of service strategies are not only based on simple things such as current data sharing, enriching resources, innovative methods, and increasing time. The more laws discovered from a large amount of data and the greater the potential value found, the faster the service level of the library will be improved. Therefore, library services in the era of big data may be more targeted and distinctive. Service methods and methods will also be adjusted with the adjustment of library service strategies. While adjusting the necessary service strategies for libraries such as literature services, information consulting, and subject services, data mining services will become the normal service content of libraries in the era of big data [5].

4. Application of data mining technology in smart library

4.1. Document information management

(1) Information interview. Traditional book information collection is mostly determined by specialized interviewers alone, or the opinions of subject experts are adopted, which inevitably carries great subjectivity and personal preferences. At the same time, the library's annual document purchase fee is limited. How to allocate between various disciplines and how to balance various document carrier forms can make the limited document purchase fund to maximize the benefits. Data mining technology solves this kind of problem. According to the analysis of the historical circulation records of the library, the corresponding relationship between the book category and the number of borrowings is drawn (see Figure 2). The book categories are compiled in accordance with the Chinese Book Classification. From Figure 2 you can clearly see the types of books that readers are interested in (I, T, H, F, K, J), so that when buying new books, the proportion of these types of books will be increased; at the same time, the amount of borrowing is small. The books are analyzed, and some outdated books and documents are removed [6]. Through the analysis of a large amount of data, interviewers can make full use of limited funds to meet the needs of readers as much as possible, increase the circulation rate of books, and provide readers with personalized services.
Figure 2. Correspondence diagram between book categories and borrowing times

(2) Book processing. Due to the large number and variety of new books purchased by the library, the book processing procedures are cumbersome, so it is difficult to put the newly purchased books on the shelves at the same time, and they need to be processed in installments and batches. In the process of book processing, data mining technology can be used to analyze the borrowing records, and the books with many borrowing times and large circulation should be processed first to improve the utilization rate of books and documents [7].

(3) Collection layout. Libraries can use the association principle of data mining and cluster analysis technology to discover the proportional relationship between various documents through the analysis of historical circulation records, grasp the law of document borrowing and returning, and determine books based on the principle of convenience for readers. The placement and order of the documents, and then the reasonable layout of the collection.

(4) Improve the enthusiasm of employees. It is generally believed that the establishment of a certain incentive mechanism can increase the enthusiasm of employees. In fact, it is not. The internal balance mechanism is the fundamental. Due to the nature of the library's work, it is necessary for us to scientifically balance the workload of employees. On the basis of the analysis results of the association principle, through the classification and cluster analysis of a large amount of data, the adjustment of the collection of documents can improve the satisfaction and work enthusiasm of internal staff.

4.2. Information Service

1) Transform the service model With the widespread application of network technology and computer technology, the library service model has changed from waiting for readers to visit to actively serve readers. The information presentation in the form of books and documents in the past has changed to the generation of information relying on new technologies. Specifically, there are the following four service modes: information retrieval service, topic setting and novelty search service, information analysis service, and personalized service [8].

2) Improve service quality In the Internet age, the role of librarians who have mastered new knowledge and new technology has changed, from a simple presenter of information to an information organizer, information servicer, information producer and information manager. Librarians use data mining technology to collect, process, process and analyze a large amount of information on readers' borrowing behavior through computer networks, determine specific reader groups, individual interests, reading habits, tendencies and needs, and use classification customization, information push and other new information [9]. The service method improves the speed at which readers can obtain the required information, thereby improving the quality of library services.

4.3. Personnel Management

1) Improve the quality of librarians Readers’ demand for libraries in the new era is not limited to simple information inquiry and feedback, but also requires librarians to provide information products with high added value. Therefore, libraries must strengthen staff management and stimulate staff innovation awareness. Based on the analysis of staff interests, qualifications and other information, a number of
distinctive professional groups have been established to stimulate the work and learning ability of librarians and provide readers with personalized services.

2) Strengthen reader management Use some techniques in data mining, such as statistical regression, logistic regression, decision tree, neural network, etc., to reveal the user’s behavior habits, hobbies, and combine the content that users are interested in with the service system to provide each user with initiative And personalized services to increase the number of library users. At the same time, data mining technology is used to analyze a large amount of data in the user database to find out the risk and reason of user loss, so as to take corresponding service measures to attract users and improve the utilization rate of books and documents [10].

5. Conclusion

Data mining technology provides technical support for the organization and management of literature resources. At the same time, it also strengthened the management of readers and improved the quality of service. The library not only has a large amount of text data, but also rich multimedia data. Therefore, librarians need to conduct in-depth research on data mining technology to guide practical work.

References

[1] Chen Chen. Intelligent library construction based on big data mining and knowledge discovery [J]. Modern Information, 2017:87-93+99.

[2] Guo Jingxia, Chen Qingmei, Li Huiping. Application of data mining technology in library construction [J]. Silicon Valley, 2012:177-178.

[3] Ha Liyuan. Construction of Archive Information Management Platform of University Library Based on Data Mining Technology [J]. Shanxi Archives, 2016:105-107.

[4] Han Qiu. Research on Library Intelligent Service System Based on Big Data Technology and Information Mining [J]. Western Region Library Forum, 2018:8-11.

[5] Kang Na, Yu Qi, Li Lin, He Qiang. Research on library intelligence service system based on data mining [J]. Library Industry, 2019:5-7+19.

[6] LU Fang. Research on Digital Library Retrieval Technology Based on Data Mining [J]. Information Technology and Informatization, 2015:178-180.

[7] NI Weiyian. Research on Digital Library Information Service Based on Data Mining Technology [J]. Journal of Zhejiang Gongshang Vocational and Technical College, 2018:88-90.

[8] Qin Zonghe. Research on Library Think Tank Information Service Construction Based on Data Mining [J]. Library Journal, 2018:43-47.

[9] Shu Yu, Zhang Lili. Personalized Service of Digital Library Based on Data Mining Technology [J]. Knowledge Economy, 2015:25+27.

[10] ZHAO Chong, WANG Lida. Design of Cloud Retrieval System for Intelligent Library Based on Data Mining Technology [J]. Modern Electronic Technology, 2020:60-63.