Research on Personalized Information Service of University Library Based on Association Rules Mining

Jie Kong
China University of Labor Relations, Beijing, 100048, China

Abstract

With the vigorous development of computer technology, network technology and modern communication technology, data mining, as a derivative of the rapid development of information technology, provides technical support for the effective management of digital knowledge resources. Today, with the digital construction of university library, personalized service of library has become an urgent need. One of the key points of personalized service in university library is how to obtain readers' personalized information needs effectively and automatically, so as to guide personalized service. The emergence of data mining technology enables us to extract the required information from a large amount of information. Association rules in data mining are to discover the connections between different items in the database, and these rules reflect the reading behavior patterns of readers. Based on the relevant principles of data mining technology, this paper constructs a library personalized information service system, and introduces the realization ways of library personalized information service based on data mining technology from the perspective of association rules.

Keywords*: Data mining, Association rules, Library, Individualization, Information services*

Ⅰ. Introduction

As the center of information resource collection, processing and service, the purpose of library is to provide good services for readers, help readers obtain and use information, and produce knowledge [1]. With the rapid development of information network technology, university libraries are successively carrying out digital construction, increasing library service mode, improving service quality and service efficiency [2]. In this process, library personalized information service has been paid more and more attention. When readers come to the library, they are often inundated by a vast sea of materials, and it is difficult to obtain books efficiently. However, the library is not clear about the needs of readers, so it is difficult to carry out personalized services according to the needs of readers [3]. At present, various types of libraries generally use the library management automation system, which will produce a large number of book circulation data every day. These data are mainly used in the daily management of the library. It is difficult to find the hidden relationships and rules in the data, and it is difficult to predict the information needs of readers [4]. There are many important information behind the seemingly disordered data. The library naturally hopes to analyze it in many aspects in order to make better use of it and guide the actual work [5]. Personalized information service is the direct embodiment of the library's "people-oriented" value concept. It is an active or intelligent push service mode to meet the needs of readers to the greatest extent, that is, through the analysis of users' borrowing habits and personality characteristics, it can actively provide users with information they may be interested in [6].

In the library, a large number of historical data have been accumulated, and many important information is hidden behind these data. People hope to carry out higher-level analysis in order to better serve readers [7]. Many libraries are using data warehouse and data mining technology to analyze book circulation data, get some potential and implicit information, and provide important reference data for libraries to improve services [8]. The emergence of data mining technology enables us to extract the required information from a large amount of information. Data mining is a non trivial process of extracting implicit, unknown and potentially useful information from a large number of incomplete, noisy, fuzzy and random practical application data [9]. Association rules in data mining are to find the relationship between different items in the database. These rules reflect readers' reading behavior patterns

[10]. Using association rule technology to mine the transaction data generated in the process of library service can comprehensively obtain the information demand patterns of various types of readers [11]. Personalized service system can provide accurate service for readers by learning these demand model knowledge. Starting from the relevant principles of data mining technology, this paper constructs the library personalized information service system, and introduces the implementation of library personalized information service based on data mining technology from the perspective of association rules.

Ⅱ. Library Personalized Service

2,.1 Personalized service analysis

Library personalized service refers to an integrated information service in the digital information environment, which mainly uses the network and information technology to obtain and analyze the information usage habits, preferences, background and requirements of each user, so as to provide users with an integrated information service that fully meets their individual information needs. To improve the construction of information resources, we need to collect information resources in many aspects and improve the quality of personalized service, so we need to have a more accurate and comprehensive understanding of readers. Personalized recommendation service, such as information push service, is a technology that automatically sends the selected books to readers according to the time specified by readers. Its essence is active service. It hardly needs readers to do anything. The system automatically provides corresponding services according to readers' information needs. An important feature of library personalized service is active service. Under the condition of accurately understanding the needs of readers, it actively promotes various library services to readers and changes the passive service mode of traditional library. The allocation of traditional library resources and the provision of services are inevitably affected by the collector's knowledge level, knowledge structure and personal hobbies due to the influence of manually collected information.

Personalized active service will enable users to obtain the best possible service with as little effort as possible. The personalized information service of digital library is to create a personalized information resource database for readers, that is, personal database. Through the library network, users submit customized personalized information to the library system for the construction of personal information database. As an auxiliary teaching and scientific research unit, on the one hand, the university library should constantly enrich various collection resources, on the other hand, it should actively improve the service quality and implement the digital construction. In the era of digital library, to advocate library personalized service, we should fully consider the application of various existing computer software and hardware technologies [12]. In terms of hardware, large capacity disk storage and parallel computing processors are becoming cheaper and cheaper, making it possible for the library to store massive resources and transaction computing. In terms of software, the progress of data mining and machine learning technology provides support for the library to provide higher quality services. The content mining based on electronic resources is to find meaningful knowledge through pattern recognition, analysis and understanding of library website information. According to the information needs of a certain field, automatically capture, collect and sort out the information required by the field, filter useless redundant information, directly provide the knowledge found by data mining to readers through information push, and actively provide personalized services.

A. Library Personalized Service System

Library personalized information service covers the whole process from website registration to receiving services provided by the system. This system first collects library user information, then builds a model of user demand behavior by correlation technology, and then compares the existing library resource knowledge base with the user model. Book resources in libraries are explicit knowledge existing in libraries, that is, knowledge types recorded on certain material carriers. Using data mining method, the relationship between readers' characteristics and information needs is mined from the library affairs database, that is, empirical knowledge, that is, what kind of readers have what kind of information needs, and stored in the knowledge base. The empirical knowledge here is the

frequent patterns hidden in a large amount of data, reflecting the tendentious information needs of different readers, so the system adopts the data mining method of association rules. Readers' behavior information is to solve practical problems and meet academic, scientific research and production needs. They reflect readers' individualized and diversified demands for digital resources and their utilization rules. Mining these data can be used to analyze and predict the trends of readers' borrowing books and returning books in the future, and put forward management decisions and layout strategies of books on this basis, which can provide reliable basis for improving the service level of libraries.

Ⅲ. Design of Library Personalized Information Service System Based on Association Rule Mining

In the current era of generalized dissemination of information resources, library users' information needs are not limited to the previous information retrieval and bibliographic inquiry, but they want to obtain deeper text information or more comprehensive answers to inquiry questions. The operation mode and service mode of library are somewhat similar to those of many commercial fields. Therefore, we should learn from the successful experience of data mining in commercial fields and apply it to library management. The library database involves a large amount of data information. In the face of these massive data, there will inevitably be lengthy and even wrong data. Therefore, in data mining, we should choose the appropriate mining type and algorithm according to the different data mining tasks, and correct, process and process the wrong data [13]. According to these characteristics of university library service objects, association rules mining method can be used to analyze the group and individual of library affairs database, so as to provide targeted group service and individual service, and change traditional passive service into active service. Using data mining technology can not only keep the integrity and functionality of information, but also deeply analyze the relationship between information, so as to meet the deep information needs of users.

In the process of data mining in library, once the theme of data mining is determined, it is necessary to define the theme first, and make clear the research requirements and purposes. Then, the explicit data and implicit data in the reader database are collected and extracted, and the related characteristics of the requirements are summarized by describing their concepts. Through data analysis, different demand classification models are formed according to similarities and differences, and the data are put into different classifications. Through the combination of demand classification model and reader behavior information, the difference analysis and deviation detection are carried out, and a large number of irrelevant data are excluded to form mining results. The data mining process is shown in Figure 1.

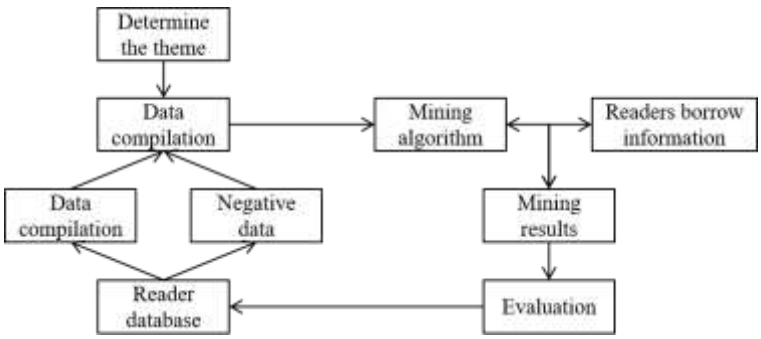

Fig.1 Data mining process

Using big data system development technology to innovate the design of precision knowledge service system, optimize and develop the precision knowledge service function and service interface of digital library, so that the precision knowledge service system can provide more functions. The prediction of user interest can be calculated as follows:

$$P_{u,i} = \bar{R}_u + \frac{\sum_{m=1}^{n}\left(R_{m,i} - \bar{R}_m\right) \times sim(u,m)}{\sum_{m=1}^{n} sim(u,m)} \quad (1)$$

Where $\bar{R}_u$ is the average score of user u on the resource, $R_{m,i}$ is the score of user m on item i, $\bar{R}_m$ is the average score of user m on resources, and $sim(u,m)$ is the similarity between users u and m.

System data collection is realized by collecting the information of readers using the library, such as borrowing information and personal identity information retained by readers swiping their cards. After the collection is completed, the resource processing layer classifies the collected information, updates the knowledge database information, and ensures the instantaneity of data resources. Individual analysis means that the system extracts specific reader records from the library affairs database and carries out association mining. Through individual analysis, readers' personalized information preferences are obtained. As individual service experience knowledge, combined with the service experience knowledge of related groups to which individuals belong, matching resources are retrieved from the resource pool according to the current needs of readers, providing readers with more accurate services. Figure 2 shows the construction structure of library digital resources under the information ecology.

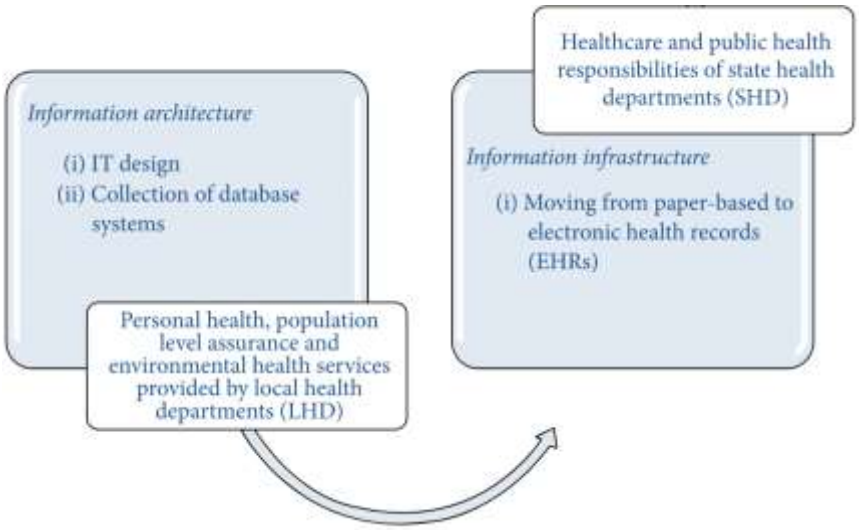

Fig.2 Library's digital resource construction structure under the information ecology

In the process of borrowing library resources, readers will leave and produce a lot of information, such as readers' personal information, borrowing history, return time, etc. Extracting information such as reader classification and subject classification from these information can provide readers with more valuable consultation and personalized information services. In addition, we can find some valuable information related to each subject and optimizing the collection layout of the library from these data. Importing and sorting data is the premise of data mining, and it is also a necessary condition for users to obtain high-quality information services. The quality of data mining is directly proportional to the mining scope and workload. In order to provide users with comprehensive and in-depth literature information services, we should fully understand the problems set by users before data mining, and expand the mining scope and depth of literature data as much as possible [14]. In the background of personalized information system, the user usage model is established according to the entered user information. Finally, the information system background uses data mining technology to obtain the related resources in the library resource pool and recommend them to users.

The transaction data of library refers to the access records of each resource pool, which needs to be integrated together to form a complete transaction database. From the perspective of integrity and security, the reader database should also be copied to the transaction database, and a table name different from the transaction record should be taken, so as to avoid directly accessing the reader database of the library system when mining reader information. Table 1 gives the accuracy results of data sets under different feature dimensions. The change trend of accuracy under different feature dimensions is shown in Figure 3.

Table 1 Results of informatization experiments based on information ecology

Feature dimension	20	40	60	80	100
Accuracy (%)	78.59	77.45	77.86	82.54	71.94

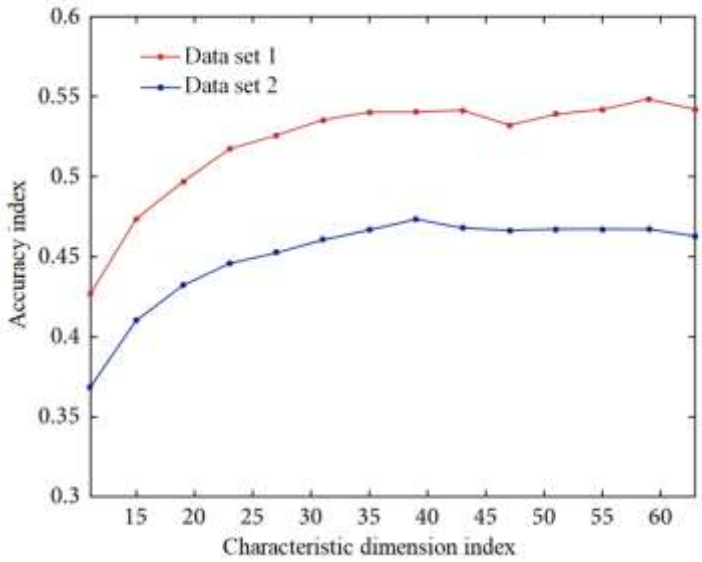

Fig.3 Trends in the construction of informatization construction patterns under different characteristic dimensions

Confidence is the ratio of the number of transactions with all items in the rule to the number of transactions with only items in the condition. Not all strong association rules are valuable to users, but the best rules are generated by calculating the promotion degree. A data mining algorithm corresponds to a model construction mode. To maximize the data mining function, the key is to choose a mining algorithm suitable for analyzing the model. After the user model is established, the model can be properly debugged according to the requirements of the system. Digital library should not only discover and absorb the hidden knowledge or patterns in information resources through information analysis and reorganization, but also describe, link and organize knowledge content and structure according to users' needs and knowledge application environment, so as to help users build personalized and dynamic knowledge maps and knowledge organization systems according to their own needs.

Ⅳ. Conclusions

In today's society, library services are no longer limited to providing simple literature search and borrowing, but instead turn to providing users with deeper and related information sources. With the development of information technology and digital library, the library, as an academic, scientific research and service institution, plays an

important role. It is urgent to process information by using classification, clustering, association rules and other technical means to promote the overall progress of library business and management. In the digital library in the future, people should explore information in the digital library with huge information resources, and provide effective decision-making basis for personalized service, and data mining tools are also indispensable. The application of data mining technology effectively solves the problem of mismatch between supply and demand in library information services, and enables readers to enjoy more effective, targeted and personalized information services. However, data mining technology involves many data processing skills and the workload is complex, so its popularity in the construction of personalized information service system of libraries in China is still very low. With the expansion of database, the application of network technology in library and the deepening of personalized service concept, data mining technology will have a positive impact on the construction of library. Facing the impetus of personalized service, the library will invest more efforts in the research and development of data mining tools.

Ⅴ. Acknowledgment

Fund Project: China University of Labor Relations, “the Fundamental Research Funds for the Central Universities”, project number (21ZYG001); Research on Accurate Servive System of University Library Resources Driven by Data.

References

[1] Ahluwalia M V, Gangopadhyay A, Chen Z, et al. Target-Based, Privacy Preserving, and Incremental Association Rule Mining. IEEE Transactions on Services Computing, vol. 10, no. 4, pp. 633-645, 2017.

[2] Wang Lei, Gao Xiang, Ma Bo. Analysis on the Path of Lending Behavior Big Data Applied in University Library Service Innovation. Journal of University Library and Information Science, vol. 182, no. 6, pp. 109-112+122, 2020.

[3] Yu Haiyang. The application of the algorithm for quickly mining the largest frequent itemsets in library management. Computer and Modernization, vol. 298, no. 6, pp. 72-76, 2020.

[4] Li Xin. Research on the Application of Strong Association Rules Mining in the Personalized Push Service of Smart Library. Information Science, vol. 36, no. 4, pp. 95-99, 2018.

[5] Chen Haijun. Design of Information Recommendation Book Management System Based on Apriori Data Mining Algorithm. Modern Electronic Technology, vol. 42, no. 23, pp. 115-119+124, 2019.

[6] Zhao Yuwei. Research on Differentiated Services of University Libraries Based on Data Mining to Perceive Readers' Needs. Library Work and Research, vol. 269, no. 7, pp. 70-75, 2018.

[7] Wei Xin. Mining Association Rules of Library Book Circulation Data. Modernization of Education, vol. 6, no. 91, pp. 133-134, 2019.

[8] Chen Shuying, Xu Jianying, Liu Yuwei, et al. The book recommendation service of university libraries under the application of association rules. Library Forum, vol. 38, no. 2, pp. 97-102, 2018.

[9] Wang Zhongyu, Liu Kaili. Book recommendation system based on collaborative filtering and association analysis. Modern Business Industry, vol. 40, no. 35, pp. 209-211, 2019.

[10] Wang Qinghua. Research on Library Personalized Fast Recommendation Algorithm Based on Data Mining Technology. Modern Electronic Technology, vol. 42, no. 5, pp. 157-159+164, 2019.

[11] Suzuki S, Stern D, Manzocchi T. The Use of Association-Rule Mining and High-Dimensional Visualization To Explore the Impact of Geological Features on Dynamic-Flow Behavior. Spe Journal, vol. 21, no. 6, pp. 1996-2009, 2016.

[12] Rekik R, Kallel I, Casillas J, et al. Assessing web sites quality: A systematic literature review by text and association rules mining. International Journal of Information Management, vol. 38, no. 1, pp. 201-216 , 2018.

[13] Qiao Ya, Wu Lin. Research on Library Smart Service Model Based on Collaborative Filtering Recommendation Algorithm. Microcomputer Application, vol. 35, no. 11, pp. 150-153, 2019.

[14] Liu Peiming, Luo Xinquan. Improvement and design of related bibliographic search recommendation methods in digital libraries. Modern Electronic Technology, vol. 40, no. 14, pp. 72-74, 2017.