

Application of Data Mining Technology in Information Analysis of University Library Literature

Gemin Li

Library, Jilin Agricultural University, Changchun, 130118

308420502@qq.com

Keywords: Data mining; Library literature; University; Application

Abstract. Data mining technology plays an important role in the field of retrieving information. It is an information processing technology in the era of science and technology, and has been widely applied in many fields. Similarly, to meet the requirements of science and technology, university library moves towards the information of network technology gradually. Likewise, Data mining technology has penetrated into information management and analysis for the university library literature little by little. Personalized needs for teachers and students could be provided by university library using the data mining technology that makes the distinguishing features of digital library play out. The author mainly studies the concrete content of the data mining technology and its application in information analysis of the university library literature.

Introduction

In the age of information explosion, the rapid development of science and technology will inevitably put forward a higher demand for data processing. People don't have no time to retrieve so huge data, but are more concerned about how to use convenient means to process date, which automatically help us to find worthy information quickly. For example, in colleges and universities, the library record of being borrowed book is a strong evidence of the using of library resources, recording the reader's personal information, and effectively reflecting the reader's interest and concerned content. The administrators of library can use data mining technology to integrate all the information from readers, and then analyze them, so that the library can provide readers with the required information services benefit to the purchase of books and the direction of upgrading the data resources, to improve the utilization rate of library information.

The Definition of Data Mining

Data mining technology is a process of discovering new knowledge using the means of network technology to find out valuable information that are hidden and cannot be found directly from the massive, random and no regular information in the data, to help people do effective analysis. In the database, data mining technology is an indispensable step, similar to the process of mining in the ore, using a special algorithm to extract useful information from the raw data. The raw data can be either structured (data in a relational database) or semi-structured (graphics, images or text). Mining knowledge can be done by adopting mathematical or non-mathematical method, deductive or inductive method. This mined information can be used to decision making, process control, information processing or data maintenance, etc. Data mining itself is a comprehensive technical process including artificial intelligence, visualization and other technical means.

Data Mining Technology

Data Mining Process. Firstly, data preprocessing. Sometimes the data mining object is chaotic, noisy or fragmented, but the quality of information can be improved after the preprocessing. Data preprocessing includes data cleansing (to fill loopholes or heal noisy data, etc.), data integration (to integrate data from different sources, for avoiding redundant data and improving the efficiency of mining techniques), data transformation (to format data), and data reduction (to minimize the

number or scale of mining objects by various means).

Secondly, data mining. After data preprocessing, the data is no longer chaotic. Then we begin to determine the purpose of work, for example, make data cluster, classify, and so on, then select the appropriate mining algorithm. In the choice of mining computing formula, people need to consider selecting different algorithms based on different data characteristics, according to the needs of users as well. Some users just want to obtain the readily understandable information, while some others want to get high-precision information.

Thirdly, the interpretation and evaluation of the results. The information obtained by data mining technology may be redundant after the evaluation and interpretation of the results, so it needs to be deleted. Additionally, sometimes the results may not meet the user's needs, then it is necessary to re-return to the initial mining phase and re-select the data mining algorithm. The factors influencing data mining include that whether the technology is pertinent or not, whether it is effective, the quality and quantity of mining objects besides. In the mining process, once the wrong data appears or undergo an inappropriate conversion, it will affect the mining results.

Data mining tasks

Several Data Mining Tasks are Described in the Following Sector. Firstly, the association analysis. If there is a certain relationship between two or more data, then the association rules can be established based on the relationship. The association between two events can be discovered using the methods of dating mining in the database. If there is a relationship between two or multiple events, one of the attribute values can be derived from other attribute values. For example, in the supermarket, people who buy milk, 90 percent of them will buy bread, so bread and milk have the relevance and both of them can be bundled sales. Filtered by the mining data, the valuable association rules can be reached.

Secondly, cluster analysis. Clustering is to find the similarity of different data, and then classify the data, where the similar date will divide into the same category. A macro concept is established on account of the application of cluster analysis, and the possible links between the data might be found. Specific clustering methods include statistical analysis (formed on distance-based classification of individuals), machine learning (no tutor learning), and neural networks (unsupervised learning, eg ART model, Kohonen model).

Thirdly, classification. Classification is to use a kind of description to structure model. The main content descriptions of classification are feature description and the distinguishing description; therein the former is aimed at the same types of objects with the same characteristics, while the latter is aimed at different types of objects with different characteristics. Firstly, the data information is fully analyzed. By analyzing the features of data in the results and utilizing data mining method appropriately, each category is found an appropriate description. Then the data to be tested are classified according to the same or different characteristics.

The Application of Data Mining Technology in the Information Analysis of University Literature

The prior style for traditional university library is borrowing and returning. However, with the development of science and technology, especially the application of data mining technology in the university library has expanded the service scope of the library. Materials in university library generally have strong professional knowledge, focusing on knowledge, covering both the latest books and a lot of their own school's diachronic collection. Due to the different levels of teachers and students in academic qualifications, so the requirements for digitization of library are higher.

Firstly, data mining technology provides the basis on library collections and procurement for university. The library for each university is the largest and the best place, where it is the soul of colleges and universities. Serving as the supporter of spreading knowledge, in order to make a progress, University needs to change the library pattern according to the requirements of the development of the times, and inject fresh blood continually. The traditional way of purchasing books is either recommended teachers or staff by colleges or directly recommended books and librarians based on experience. This traditional way of buying is subjective. In order to improve the

relevance of book purchased and improve resource utilization, data mining technology is essential to be applied to count readers' borrowing records, study utilization rate of different types of books, integrate literature and related data, and provide the basis for book procurement staff to purchase targeted books according to the relevance of different disciplines.

Secondly, the application of library literature and information. The retrieving department are set in the entrance of University library, where teachers and students could acquire targeted book information by entering the keyword. While because of the large number of general library collections, management personnel could timely discover books in the state high or low utilization rate through data mining technology, thus adjusting the book structure by increasing or eliminating the worthless books timely. Using data mining technology, we can excavate all the readers' professional knowledge and record from the entire database of the library, so as to know the professional and interest of the reader and then push the popular data in time. This will help the user narrow the retrieval scope of interest, for example, the number of DELPHI and database construction classes are usually borrowed at the same time, so that library managers can store these two books together, pushing services initiatively, which can improve usage rate effectively, and providing the current and potential users with personalized services. Similarly, the application of data mining technology can find hot and popular books in time, to further optimize the distribution of library collections.

Thirdly, data mining technology can be recommended in the following circumstances. First of all, when readers browse the retrieval system, sometimes many students borrow some books but do not know what type of books for themselves, so as to browse the main interface aimlessly. In that case, the library can use data mining technology to release the best-selling number and publish the ranking list of reading, etc. That makes readers understand the latest books and recommends them, having the association in the views number, in a bundled mode. Second, when landing the public retrieval system, once readers visit the public retrieval system, mining technology can identify the readers' borrowing situations through the user's account. Using the principle of clustering to find the number of readers who have the similar interest, retrieval system can recommend targeted books to readers. Besides, when the user enters into library, according to data mining technology statistics based on the browsing bate of readers, the books with low browsing rate can be placed on the main library, while the latest books are on the open-shelf stacks. In addition, the library can also issue the customized magazine of library to users at regular intervals, where some introduction, comments and the number of recommendations of new books are published on it. When readers borrow books, system could make them leave Email, and regularly push the latest materials and information. Finally, the library can use the data mining technology to count the readers' borrowing history after the readers finish borrowing, and recommend related bibliography to them according to the calculation results of borrowing data. Even through readers don't come to the library, library could also push them possibly concerned books via Email.

Fourthly, to avoid the invasion of the library database. Database is the core of the library, there are the collections of data, as well as the student personal information and borrowing history. If the database has been invaded, the consequences are serious, even may return to the traditional model. In order to ensure the security of the database, the firewall can be installed on the server, and set secret security. In addition, using data mining technology to set up abnormal access barriers, through the sequence of normally frequent events, we confirm the access outliers and organize behavior of destruction timely, then report the behavior of destruction to avoid damaging the data.

Conclusions

With the development of science and technology, the knowledge and information is blowing, so the library needs to be innovated continually. Data mining technology is applied more widespread in the analysis of library literature information. It is of great significance to improve the service level of digital library and optimize the library bibliography arrangement, provide college teachers and students with the convenience as well. This paper introduces the meaning, process and important task of data mining. It can analyze the readers' information, borrowing records and retrieving

information to meet the personalized requirements of the users, and provide valuable resources and services to the users. The exploration is expected that it is helpful to the future development direction of university library and the application of data mining technology in library information analysis.

References:

- [1] Wand Cui-ping. Comparative study of Mylibrary personalized service system in domestic and foreign libraries. *Information and Documentation Work*, 2004, (367-70)
- [2] Zhang Fengqiao. College library personalized information service situation and development countermeasure under the network environment. *Agricultural Library and Information Service*, 2005,17 (8): 54-55,58
- [3] Guan Enzheng, Zhou Chunguang, Xu Xiujuan, etc. Frequent sequence pattern mining algorithm. *Journal of Jilin University*, 2005,6 (5): 267-268
- [4] Lu Feng, Zhang Weiwei. Four kinds of sequence pattern mining algorithm characteristics. *Journal of Wuhan University of Technology*, 2006,28 (2): 57 -59
- [5] Mai Yonghao. Data Warehouse and data Mining methods and their application in public security information construction. Doctoral Dissertation, 2000
- [6] Zhou Pei. Data mining technology in the library system application research. Southeast University, Master of Engineering thesis, 2006
- [7] Deng Xiaomei. Telecommunication customer segmentation model based on data mining. Dalian University of Technology Master Thesis, 2006.
- [8] Sun Weixiang. Data mining and information fusion based on fault diagnosis method. Shanghai Jiaotong University Doctor of Engineering, 2006.
- [9] Fan Zhonglei, Pan Longfa A customer database model based on call center and data mining. *Computer Application Research*, 2002 (1), 84-85.
- [10] Chen Li, Liu Jing, etc. *Intelligent data mining and knowledge discovery*. Xi'an: Xian University of Electronic Science and Technology Press, 2006,1-5
- [11] Yan Yu. Data mining technology in college students employment guidance decision-making. *Journal of Shenyang University of Technology*, 2007,29 (3): 234-236
- [12] Li Meng, Deng Chuanguo. Association rule mining algorithm and its application in student quality evaluation. *Computer Development and Application*, 2006,20 (7): 66-88