

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(21), 2014 [12804-12809]

Analysis on approach of data mining in personalized information system of university library

Yongguang Chen

School of Educational Science, Zhoukou Normal University, Zhoukou, 466001, (CHINA)

ABSTRACT

With the development of network technology, university library construction is gradually moving in the direction of digital libraries; digital age of university libraries has arrived. By automated information management system in libraries, readers can easily inquire book information and borrow books, as well as manage bibliographic information retrieval and WEB access record query. In order to know how to analyze and predict readers' needs and saving trends, data mining technology is introduced, through which readers can easily find the useful information which can be utilized from mass of data. This paper begins with the basic concept of data mining, systematically introduces its concept and implementation methods, and then analyzes the related knowledge of personalized information system in university libraries accompanied with comparative analysis of its development status. It is found that university libraries lack the ability of data mining. Based on the circumstance, this study designed a set of model structure combined with library system in order to achieve personalized information system based on data mining technology. Then the study analyzed the composition of each module from a theoretical point of view as well as the explanation of functional description and workflow. Model building provides a better theoretical basis for the development of university libraries, thereby improves application ability of university libraries.

KEYWORDS

University library; Automated information management system; Data mining technology; Personalized information system.



INTRODUCTION

With the rapid development of computer communication technology, university libraries have entered the information age. As the center of information and intelligence searching, university library has experienced harsh information revolution. The starting point of all university libraries is reader-centered, which explains the emergence of personalized information system, the personalized information system is gradually replacing the old-fashioned library services, it changes the information processing, collection and dissemination of libraries in a large extent, optimizing service mode^[1]. personalized customization is regard as the head of seven trends of future library development internationally^[2]. However the investment of university libraries in China is not enough. Even though certain library homepages have appeared the project of “personal library”, the interiors still lack the substantive content. In order to respond to this era of growing individual needs, university libraries should carefully and accurately analyze user data, then provide readers with the discovered potential service initiatives^[3]. It requires library to be able to find the useful information from mass of data. Therefore data mining is produced and widely spread. This study analyzed the development situation of university libraries and technical competence of data mining techniques, and specifically analyzed the combination of both; established a data mining model which is suitable for university libraries.

RELATED BASIC CONCEPTS

Personalized information system in university libraries

American colleges and universities had already begun the program of digital library program in 1991, and started to provide personalized information services. Soon afterwards Europe and other developed countries began the similar research. With nearly 20 years of rapid development, foreign university libraries’ personalized information system has been carried out smoothly. The main representative is the “My Library” in Cornell University. It can periodically send periodicals, directory and other personalized services to users, which is a good solution for the problem of spam and information overload. At present, common personalized information systems at abroad are mainly as follows: Tapestry system, Fab system, Citeseer systems. Research of personalized information service system in Chinese university libraries is far behind. Only from the year of 2000 the related research emerged^[4]. Then universities like the People’s University, Zhejiang University and Wuhan University began to launch its own personalized service system. TABLE 1 displayed the comparison of personalized information system of university library in China and foreign countries, from which we can see the gap between them. At present, personalized services of library in all colleges and universities are quite simple; rich service forms are needed.

TABLE 1 : Comparison of personalized service of university libraries in China and foreign countries

University library Services	North Carolina State University	Cornell University	University of Toronto	Chinese Academy of Sciences Library	Zhejiang University	Huazhong University of Science and Technology
Customization of library resources and other WEB resources	√	√	√	√	√	√
Personal library management	√	√	√	√	√	√
Latest resource announcement management	√	√	√	√	√	√
Library verbal system						
OPAC’s						
Access	√	√	√	√	√	√
Personal favorite links	√	√	√	√	√	√
Customization of page resource layout and font color	√	√	√	√	√	√
Library catalog check and borrowing records	√	√	√	√	√	√
Providing interlibrary loan and document delivery service	√	√				√
Personalized pages and resources display	√			√	√	√
Bookmark function			√		√	
Connecting search engine	√					
Shared library			√			
Automatically selecting resources in library			√			
Update and download						

Personalized Information service center in university libraries is for single user or reader in order to provide each reader with unique information service, which requires the system to be able to analyze personal information, identify personal preferences, and predict what kind of service to provide, meanwhile optimizing users' feedback. Figure 1 shows the common structure of the personalized information system:

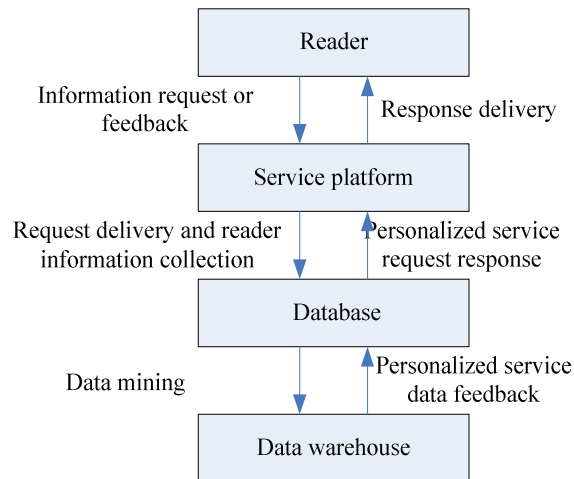


Figure 1 : Personalized information system structure

Data mining technology

Database system experienced continuous development and innovation in the 21st century, continuously improving the ability of data processing at the same time. Of course, it also brought a huge amount of data, which requires the ability to find data that can provide users effective decisions from the massive amount of information; naturally the data mining techniques emerged, and its role and platform are growing bigger and bigger^[5]. Figure 2 is the common data mining process model:

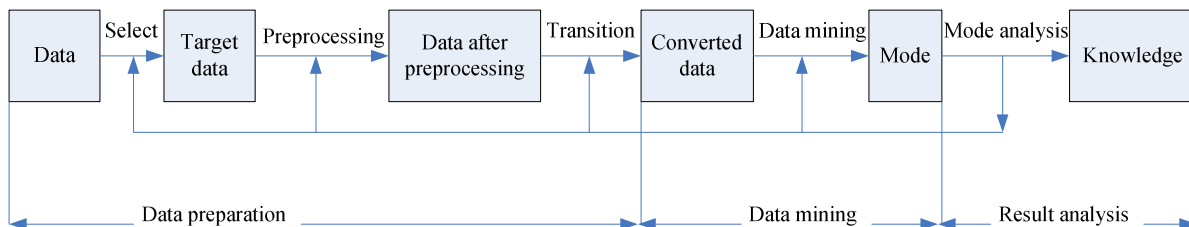


Figure 2 : Data mining process model

The difficulty of Data mining is to extract valuable information from the huge, random, and incomplete data information. Data mining technology is to build a complete model through a variety of tools and methods, and to find the rules and information. Generally data mining techniques include several methods listed below^[6]:

(1) Decision Tree: a relatively simple knowledge representation. It can reason classification rules from a set of events or data without rules and orders, and organized into the tree form for representation. It gradually expands the cases from top to bottom; forms different categories; compares and categories property value at each node; estimates downward branch, and makes conclusion in the end.

(2) Nearest neighbor clustering^[7]. It can predict when a record will appear the predicted value. The method is to obtain similar predictive record through historical data; then the predicted value is the closest recorded value in the unsorted record.

(3) Statistics: It is based on relatively sophisticated and improved mathematical statistics, well explaining the data, and analyzes data based on the function relation and correlation relation. Generally there are factor analyses, discriminant analysis and partial least squares regression.

(4) Neural networks: it is similar to the neuron of human brain; it can complete the data processing, storage and learning through a parallel way. It has already achieved clustering, forecasting and recognition in data mining of university libraries.

(5) Association rules: it is the major data mining method; the principle is to find the correlation between two things or many things. It is currently the most widely used method.

Data mining technology is based on a lot of methodologies and algorithms; it mainly includes the following categories: statistical sampling, artificial intelligence, information theory, signal processing, and modeling techniques. Data

mining technology is produced by the integration of these disciplines. Therefore it can not only analyze mass data; but also predict results. Figure 3 shows the relationship between data mining technology and other areas.

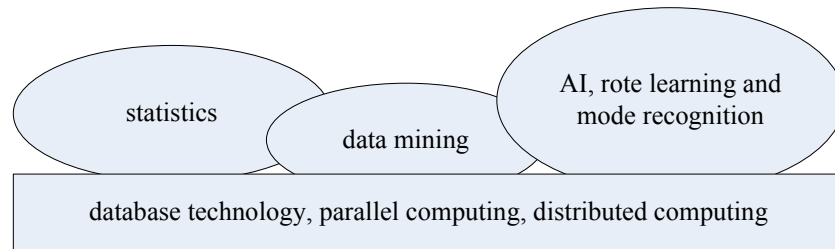


Figure 3 : Relationship between data mining technology and other fields

APPLICATION FRAMEWORK OF DATA MINING TECHNOLOGY

Construction of personalized information system in libraries

How to build a personalized information system on the existing traditional library system is the primary challenge. Figure 4 shows the current common model structure based on data mining. The first thing is to collect and organize users' essential data information; then analyze each user by personalized service model; conform users' needs, habits and interests; finally confirm users' potential needs, improving the service instruction of university libraries^[8].

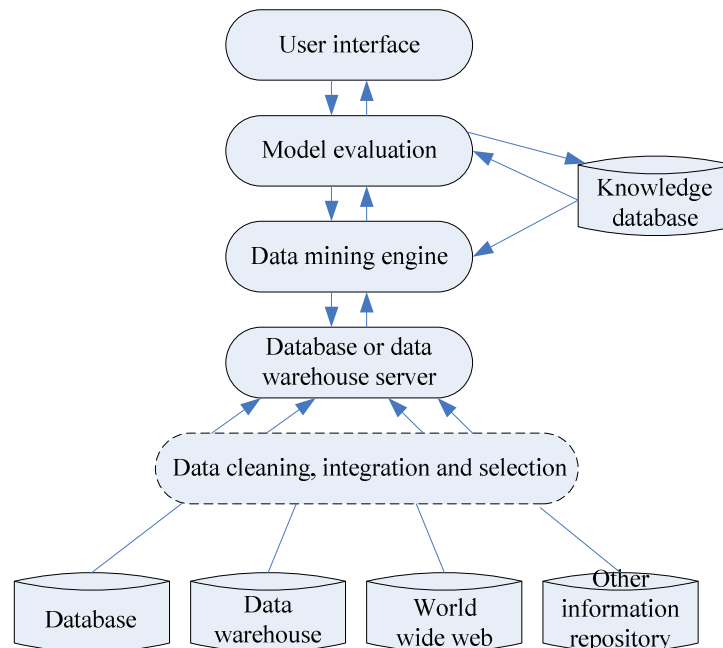


Figure 4 : Data mining system structure

At present data mining systems in university library generally need to provide the following function^[9]:

- 1) Classify reader groups; unify similar readers, so that personalized services can be made towards specific groups of readers;
- 2) Analyze library data of the entire university library, including the lending situation. Make cluster analysis on certain popular books, in order to help managers optimize library book layout and improve the utilization of the book;
- 3) Make association analysis on books, which provides the user with a list of books they might be interested, avoiding a haystack type of search process;
- 4) Conduct actual promotion and application, including bibliographic recommendation.

Implementation of data mining technology

Data mining is the primary content to provide and improve personalized service, and data collection is a prerequisite for the data mining, therefore the quality of the collected data is related to the quality of data mining information. This study comprehensively analyzed users' data records and features by the application of Clementine data mining software; identify the appropriate related information through the appropriate mining methods, so as to complete user cluster and build database

for classified results. The normal methods is online and offline mode. The offline mode is mainly used to complete the work content which needs a long time, and store the result data for online recommendation.

The key point of data mining is to establish a comprehensive data warehouse, which is the place to store the collected data. Generally the accuracy of personalized service should be ensured, so it needs to build summary tables and views like the below (which can be increased according to demand) for data storage.

```
Reader : reader_id, department, class ;
Catalogue : catalogue_id, catalogue, content ;
Book_browse: reader_id, book_id, time, browse_type, browse_content ;
Keyword : keyword_id, keyword_content ;
Book_index : book_index_id, catalogue_id ;
Single_read_preference: preference_id, reader_id, catalogue_id, preference_value;
.....
```

After the completion of the database establishment, adding content of data warehouse and analyzing the inner have become the main part of data mining. Here is an example for the added content: When a user searches books by keyword, he or she can add content in the "keyword list".

The following is the data analysis, such as predict readers' preferences by the frequently used keywords and browsed book categories, as well as loan records and so on. Normally for a single reader, the system usually analyze the possible interested contents from the "browsing history" and "borrowing records", and sort the results of the analysis from high to low, so as to establish a sound reader preference table (single_read_preference). In addition, the system can also further improve the results of data mining from "browsing history" and "loan records", which is to rank the browsing and borrowing times of certain book category. Therefore it can be intuitive to reflect the popular and recognized books. The result can also be used to build book_count_lb.

Data analysis is required to complete the data mining modules, which mainly include book clustering, reader clustering and association rules mining of books.

Basic data of book clustering is the borrowing times of books clustering, and it generally has two directions, a total number of lending and year. The total borrowing time is from all lending times after the book hit shelves; the year data is the reflection of the current year or the prior years' lending times. In this way, the book utilization and library collection can be cluster analyzed to determine the hot books. Hereby establishing hot book area and rational library collection distribution is very helpful for book utilization and book purchasing.

Cluster analysis data of reader interest is derived from everyday use situation of readers, such as certain readers go to library in a few days of every week, while some readers only go to the library right before exam in the end of year, and some readers even don't go to library through the entire semester. As a result, the same lending regulation would be very unreasonable. For example, every reader can only borrow four books every time, and the deadline is thirty days. Undifferentiated rules and conditions are lack of humanization for readers who love reading. Therefore, the university library can use bank credit card use method to better serve readers, changing the old rules. It varies with each individual to better attract readers' attention.

Association rules mining of books are to look for potential lending rules from readers. For example, the reader will intend to borrow the up and down of novel at one time, and some references will be looted at the end of the period. Some rules even cannot be accurately predicted in daily life. Then the use of association rule in books can find the laws we ignore in time. It can better find readers' lending preference, which helps better provide personalized services.

Providing users recommend research is generally through their own browsing history and lending data. Meanwhile the system finds the similar user from other saving database, completing the work of recommending books. Generally data mining implementation process is as shown in Figure 5:

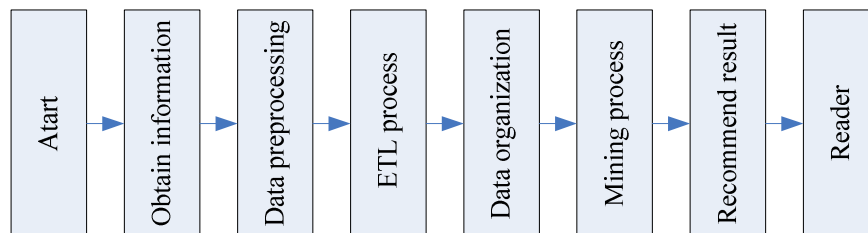


Figure 5 : Data mining implementation process

CONCLUSION

With the rapid development of computer science and other related technologies, the personalized services for university library become major trend of development, which can provide personalized information system based on the characteristics and professional background of the users. Of course, to better achieve this function, data mining technology should be applied. Conduct comparative analysis on mass data of library; build the final user information model; improve the

attendance rate of readers; and optimize the collection structure of library. Personalized information system of university library is a project of extraordinary significance. It is highly effective for enriching reader interest and upgrading library application.

REFERENCE

- [1] Tang Qihong, Cao Hongbing, Tang Xiaoxin, Li Gao, Gao Song; Study on special subject of personalized university library [J], *Research on Library Science*, (13), 53-59 (2012).
- [2] Li Jiaqing; Study on personalized information service method and strategy [J], *Modern Information*, (9), 45-48 (2006).
- [3] Zhang Fengqiao; Current situation and development strategy of personalized information service in university library under network environment [J], *Journal of Library and Information Sciences in Agriculture*, 17(8), 54-55, 58 (2015).
- [4] Liu Zuzhi; Research status of mylibrary in China [J], *Books and Intelligence*, (2), 104-106 (2006).
- [5] Zhang Jiaqin; Application of data mining in university library's personalized information service [J], *Luohe Vocational and Technical College Journal*, 12(6), 188-189 (2013).
- [6] Niu Genyi; Study on data mining in Chinese libraries [J], *Modern Information*, 29(1), 128-130, 133 (2009).
- [7] Xiong Yongjun, Chen Chunying; Personalized push service in digital library based on association mining technology [J], *Library and Intelligence Service*, 54(1), 125-129 (2010).
- [8] Zhang Lu, Tan Yuehui, Su Chunping, Wang Dong; Application of data mining technology in readers' lending behavior analysis [J], *Information Technology*, (6), 36-40 (2005).
- [9] Yang Fang; Application of data mining in university library's personalized information service [J], *Sci-Tech Information Development & Economy*, 30(21), 50-52 (2011).