Open Access

Deep Analysis of Data Mining Method in Personalized Information System of University Library

Kunpeng Wang[*]

Library China West Normal University, Nanchong 637000, China

Abstract: In order to discuss the application method and execution process of data mining in personalized information system establishment of university library, the thesis introduces existing condition of university library and insufficiency of the information service system. At the same time, data mining technology is introduced to simply describe the data mining process and introduce two top applications of the data mining technology in personalized library information system, namely student interest guidance quality and establishment of relevancy rule. Furthermore, more classical algorithms (FP-growth algorithm and K-mean clustering algorithm) are introduced in the data mining technology in detail. The data mining technology is a new data processing method. Nowadays, as for high flux reactor data analysis, data mining technology becomes more and more important in the construction process of personalized library information system.

Keywords: University library, data mining, interest guiding, relevancy rule, Fp-growth algorithm, K-mean clustering algorithm.

1. INTRODUCTION

Information Technology (IT) means modern technology related to information. There are different interpretations from different people and in different books. Network information technology has become the common technology for various industries in the 21st century, because all management activities depend on information [1]. The so-called informatization is to reconstruct other industries or disciplines with information technology, thereby improving the enterprise benefit. During the process, information technology assumes the role of the competent tool. In up-to-date information age, information technology has brought our daily life great convenience [2]. University library serves university teaching and scientific research and it means the documentary information center of a university. With progressive development of the physical science and social science, plenty of books have been introduced into university library for storage in recent years. A great deal of data information has been accumulated in the database of university library. Therefore, requirements on the personalized service and information supplied by the library are improved continuously. One of the problems confronted by and to be solved by the library personnel is how to accurately screen out the book information required by the readers among the library book resource with enormous data. In addition, how to effectively manage the library holdings is also a puzzle confronted by related personnel [3, 4]. Data mining is a step to find knowledge in the database. Generally, data mining is the process to search the hidden information from a great deal of data through algorithm. Related to computer science, data

mining realizes above objectives through statistics, online analysis treatment, information retrieval, machine learning, expert system (based on previous empirical law), pattern recognition and other methods. Data mining technology supplies a direction to the university library in the development research of personalized service. Based on different information demands from readers, libraries use data mining technology to mine and analyze a great deal of borrow information stored in the library database so as to find out the hidden book relevancy rule, reader's borrowing preference and habit. Based on these results, we can instruct personalized information recommendation work of the library and supply readers with high-grade personalized service [5, 6].

2. OVERVIEWS OF DATA MINING

Data mining is to put forward the knowledge which interests people from the large database. However, implicit and previously unknown, the knowledge has potential value in decision; the abstracted knowledge can be expressed in concept, rule and other forms. These rules contain special relation among a group of objects in the database and reveal certain useful information. Therefore, they can provide basis for market planning, financial prediction, operation decision, etc. Data mining process is practically a knowledge discovering process where various data processing techniques are applied to dig out potential models or rules from the assembly between a great deal of factual data and observation data and better understand the relation among data [7]. Through data mining, potential, effective, novel, valuable and finally apprehensible knowledge information can be abstracted from a large data base or related data in the data warehouse and display them at a different angle. In this way, the large database with a great deal of abundant and reliable resource can serve knowledge induction.
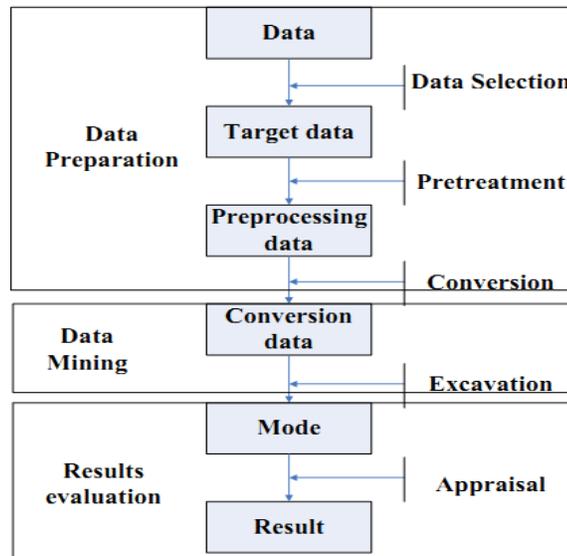
Fig. (1). Data Mining Process.

In the data selection process, extract a group of data to be operated from the source database of the mined jobs and then perform pre-processing works such as data elimination, supplement, etc. After pre-processing, we can obtain safe data applicable to mining operation. After format conversion, first-phase preparation for security data is completed.

Data are classified or clustered based on similarity degree among data objects so as to find association rules among these data items, seek characteristics among data and select related algorithm for mining works, thereby obtaining useful modes for job completion.

After mining works, the obtained mode result shall also be interpreted and evaluated. The whole data mining process is a continuous feedback process until a satisfactory result is obtained. The data mining process is shown in Fig. **(1)**.

3. PERSONALIZED INFORMATION SYSTEM OF UNIVERSITY LIBRARY

History of university library can be traced to middle of the 12th century when a medieval university was constructed in Europe. Since then, university libraries in countries other than China have been gradually established and they developed slowly. Paris University and Oxford University libraries are taken as the forerunner of European University libraries. In relation to western developed countries, it is only a short time for China to open a modern university library. The recognized university library of China in modern times is Imperial University of Peking (the precursor of Peking University) established in 1898. Till June 1987, there were totally 1,053 libraries attached to common full-time universities and colleges. At present, the biggest university library in China is Peking University library.

University libraries serve both teaching and research and they are one of most important bases for talent cultivation and execution of scientific research [7]. Their major jobs include collection of different types of document literatures, scientific processing sequencing and management as well as supply of literature resource guarantee for teaching and research activities of universities; execution of reading and reader instruction; performing reader education, culturing information consciousness and utilization skill of documentary information among students and teachers; performing reference inquiry and information service work, developing documentary information resource; planning and coordinating the documentary information work of a university as a whole; performing cooperation among libraries, achieving resource sharing in larger range; participating in integral construction of library and information cause all over the country; performing academic research and communication activity.

Personalized information service of university library started from 1980s. Up to now, library information service has undergone traditional information service, composite library information service and popular digital library information service. At present, many universities in China still use previous borrow mode. Computer and other modern tools are frequently used in borrow inquiry; borrow management and other simple works. Inter library borrow, technical novelty retrieval, information retrieval, subject service, etc. are passive services for visiting readers.

Personalized service of the library is to provide users with information resource and function to meet their personalized requirement based on their use behavior, habit, interest characteristic and special demands. It is to fully show reader's first concept.

At present, most universities in China still use previous traditional borrow modes and the computer technology is mainly used in auxiliary borrow and book management. During the borrow process of books from library and the

reference process of electronic journals, a great deal of information is left and generated. Under the informatization condition, the information is stored in the database in recording form. Digital library knowledge and information are to combine information service object, information resource and information technology to continuously collect user data such as users' index structures, reading interest, various stock number and ratio, library utilization frequency and mode, service demand level and satisfaction degree, probable development change and other parameters so as to deeply study user information demand, establish clear and ordered user information feedback channel, scientific, feasible and systematic evaluating indices, accurately reflect and evaluate service operation status and efficiency of libraries.

Personalized information service also pays special attention to expanding knowledge intension and category, realizing knowledge mining and knowledge finding; extending knowledge covering surface of related problems as far as possible; taking full advantage of existing library entity and virtual network resource; supplying the library with informatization service in more extensive surface based on modern information technology.

3. APPLICATION OF DATA MINING IN PERSONALIZED SERVICE OF LIBRARY

In the borrow database of a library, there are a great deal of data information related to readers, including student ID, name, grade, information of borrowed books, borrowing history condition, etc. Mathematical statistics at different layers, mathematical analysis and related cluster analysis can be performed with different data attribute information. In accordance with cluster analysis, various typical clustering groups can be obtained based on cluster analysis. Here, we take students as the subject investigated to divide readers into different interest groups. Therefore, we can divide students of different ages and grades in a university into different interest groups and then perform relevant mining. Classical clustering analysis algorithm contains partitioning clustering method, level clustering method, density based clustering method, model based clustering method, etc. We will simply analyze the partition based clustering method with K-mean algorithm below.

K-means algorithm is a clustering algorithm and the main purpose thereof is to divide the n object into k categories based on their attributes (values) and rules. At this time, it is required that k < n. The clustering principle of the method is to partition K category and find natural cluster centre among data. It is supposed that we have totally n data points to be divided into k clusters. The objective function to be optimized by the method:

Where, r_{nk} is 1 when it is classified into cluster K. Otherwise, it is 0. To minimize the objective function, K-mean algorithm is realized through alternative method. Specific steps to realize K-mean algorithm:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \, \mathrm{P} x_n - \mu_k \, \mathrm{P}^2 \tag{1}$$

1) Select K at random from data N as the center of K clusters;

2) Calculate the similarity degree of each element in N with K-center and classify various elements into corresponding categories based in minimal similarity degree;

3) Calculate different centers again based on category 2 result;

4) Repeat 2 and 3 until an optimal result is obtained.

Mining of relevancy rule is to find hidden interdependency in large databases so as to help people perform administration and make more accurate decision. In relevancy rules, common algorithm: FP-growth algorithm, Apriori algorithm, partition based algorithm, etc. We shall perform simple analysis for FP-growth algorithm below.

FP-growth algorithm is to compress transaction database into a FP-tree and then seek relevancy rule through traversal of the tree.

We will use examples to illustrate FP-growth algorithm below: 9 customers buy 5 commodities A, B, C, D and E and the buying condition is shown in Table **1** below.

Table 1.   Data sheet for purchase of commodity by customer.

Customer	Commodity	Customer	Commodity	Customer	Commodity
1	BAE	4	BAD	7	AC
2	BD	5	AC	8	BACE
3	BC	6	BC	9	BAC

It is not difficult for us to observe from the table above that the purchase support degree from customer on B commodity is 7 and support degrees on other commodities are A:6, C:6, D:2 and E:2 respectively.

Null node is established for the constructed FP-tree as the root node to scan the first affair and construct the first branch of the FP-tree; scan the second affair and the first item of the 2nd affair is B.1 is added to the existing node. Next item is D and there is no this branch under B. Therefore, a node is constructed as a sub-node of B. At the same time, the support degree is set as 1. Scan the third one and the first item of the third affair is B. 1 is added to the support degree of the existing node. Next item is C and there is no this branch under B. Therefore, a node is established as one sub-node of B. At the same time, support degree is set as 1. Scan the fourth affair and the first item of the fourth affair is B. 1 is added to the existing node. Next item is A. 1 is added to the existing node. Next item is A. 1 is added to the existing node. Next item is D and there is no this branch under A. Therefore, a new node is constructed as one sub-node of A. At the same time, the support degree is set as 1. Affairs are scanned in the node and 1 is added to the support degree of existing node. Add a new node if there is no branch until the ninth affair is scanned. Structure of FP-tree is shown in Fig. **(2)**.
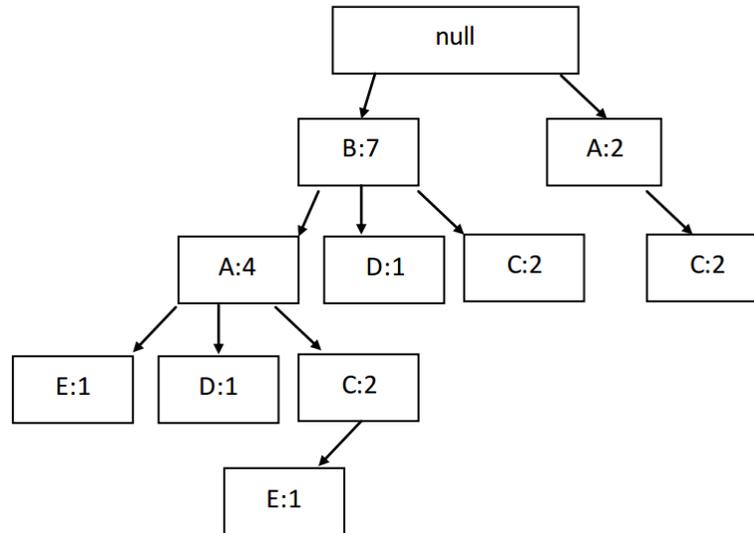
Fig. (2). Structure of FP-tree.

Find frequency item cluster based on constructed FP-tree. The algorithm is to find condition mode base one by one in an ascending sequence of the support degree based on frequency item cluster. E is taken as the example and two common branches of E (B, A, E:1) and (B, A, C, E: 1) appear in the frequency mode tree. They both take E as the suffix. Their prefix paths (B, A:1) and (B, A, C:1) compose condition mode group of E. Based on these two condition mode bases, frequency mode (B, A, E:2),(B, E:2) and (A, E:2) are composed. Support degree of assembly (C, E:1) is 1. Delete the assembly if it is less than the minimal support degree. Finally, we obtain a frequency item cluster table. In the end, relevancy rule can be mined based on frequency item group.

CONCLUSION

In the digital times with explosion of knowledge, knowledge study, information searching, information acquisition mode become diversified gradually. At the same time, in the numerous and jumbled knowledge system, it is important to learn how to accurately obtain useful information. There are abundant paper and digital resources in university library. In face of multifarious data information, it is difficult for people to obtain required resource content at a simplest and most effective mode. Therefore, in recent years, how to accurately and quickly retrieve library information by undergraduates has become the hot topic in the education of present undergraduate. Presently, related course is established in many universities to make up insufficient borrow system of old library. However, no source is traced and the problem is not solved from the root. Nowadays, at high flux data processing of university library, data mining technology has pioneered a new route to construct a personalized information system for university library. With students the subject, the data mining technology is to seek intrinsic contact and rule from a great deal of borrow database of university library, induce and classify cluster at back of data through deep processing, mining and aggregate analysis of data so as to provide practically valuable management clue for informatization service system of the library. Furthermore, the data mining technology can also provide multi-level and intelligent information and knowledge service system. The technology becomes mature continuously and it provides core technical support for informatization construction of the library in the future. At the same time, the method can greatly promote library service level and interacademic status

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1]   F. Yang, "Application of data mining in personalized information service of university library", *Scientific and Technical Information Development and Economy,* vol. 21, no. 30, pp. 50-52, 2011.

[2]   M. Fang, "Overview of personalized service of university digital library", *Technical Innovation News,* vol. 33, pp. 219-220, 2010.

[3]   X. X. Liu, "Study of recommended method for personalized information data mining of university library", M.S. thesis, *Liaoning: Liaoning University,* 2013.

[4]   H. X. Gui and X. R. Meng, "Research of mining algorithm for efficient relevancy rule based on Apriori", *Journal of Anhui University of Science & Technology (Natural Science Edition),* vol. 29, no. 4, pp. 55-58, 2009.

[5]   X. Y. You and Q. Yang, "Design and application of a network based intelligent and personalized learning system", *Computer Engineering and Science*, vol. 28, no. 7, pp. 103-104, 2006.

[6]   Y. H. Lv, J. L. Chen, H. Jin, and B. Y. Ruan, "Application of data mining in university library", *Scientific and Technical Information Development and Economy,* vol. 17, no. 18, pp. 3-5, 2007.

[7]   J. Li, "Application research of data mining technology in personalized service of university library", *Tianjin: Tianjin University,* 2012.