Application of Data Mining Technology in Book Partition System Design

Kui Huang

Department of Library, Henan Vocational College of Water Conservancy and Environment, Zhengzhou 450008, China

Abstract

In this paper, we introduce the algorithms of classification, clustering, recommendation and so on, and then compare and analyze the performance and the scope of the algorithms. Finally, we choose the K-means clustering algorithm and Apriori's frequent item discovery algorithm and apply them to the system. The readers clustering, book clustering, readers borrowing behavior analysis and other system functions are introduced. User-centric system is designed to distribute part of the administrator rights to individual users, so that users can also participate in the management of activities under the permission to allow both administrator time and cost savings, but also allow users to get a better sense of experience.

Keywords: Data Mining, Book Partition System, K-means clustering algorithm, Apriori's frequent item discovery algorithm.

1. INTRODUCTION

University library is a university student learning, where students can access to the temple of knowledge and work efficiency in the library. Thus it not only can save students find, search books time, but also to some extent, help students solve difficult problems, speed up the learning progress and enhance interest in learning. Based on the types of students' basic information in Henan Vocational College of Water Conservancy and Environment, this paper develops and designs a brand new intelligent digital library management system. Realization of the library books, students borrow information online query function, but also for different users personalized recommendation (Qinghua, 2016).

At present, most of the services provided by the library management system are still based on the manager, and the system design requires the managers to manage them conveniently. Such an arrangement may be more manageable. However, inadequate consideration of the user experience of using the system tends to make the business logic more cumbersome and the user operations complicated (Wang et al., 2011). In this paper, the system designed will make full and effective use of the relevant documents generated by the readers when they use the library management system, such as readers 'search logs, readers' historical borrowing information, download records and other personal data to fully exploit the user data from the user from the point of view, like user-recommended streamlined data and bibliography (Zhong et al., 2010). Personalized services not only be able to take the initiative to track user habits to provide personalized service, but also pay attention to service quality, which allowing users to get the best experience (Wu et al., 2012). The central goal of personalized service and the point that needs the most attention are the degree of satisfaction of the users. Facts have been proved that the application of data mining technology library management system and personalized service are of practical strategic value and research significance. As shown in Figure 1, the example of the book partition system is given in details.

The starting point of this paper is to explore the basic concepts and related technical details of the existing data mining to apply the technology to the library management system in an appropriate way. So that it contains the main collection and analysis of user data, mining data potential association, and predict user behavior (Liu et al., 2011). Specifically, according to the details of the relevant parameters in the campus library system of Henan Vocational College of Water Conservancy and Environment, a more intelligent and more digitized personalized recommendation library is designed and developed so as to improve service quality and management efficiency and establish a perfect personalized E-library to improve user experience, which can also improve service quality and speed up the construction of our personalized library information system.
2. BOOK CLASSIFICATION SYSTEM BASED ON DATA MINING TECHNOLOGY

Data accumulation in the database, transmission channel, results in a large amount of waste of resources. This makes data mining technology has a very wide range of application prospects (Zhou et al., 2014). In this topic, data mining is from the database of digital library system, which using the user's historical search records, historical borrowing records, areas of interest and other data, tap the other potential needs of users. As shown in Figure 2, the data mining technology is given in details.

2.1 Association rule analysis

Association rules analysis is to find an event and other events between dependent or association, and the association rules of association is a technical analysis after abstraction. It can be found that a large number of seemingly disparate, disorganized data, potential links deep (Wu et al., 2012). The name of association rule was first put forward by R. Agrawal et al. It is one of the important techniques of data mining by mining data to solve the problem of transaction database analysis. Association rule mining is to find the potential association rules between data items or attributes and variables in the database, and the famous examples are beer and diapers". Association rule mining firstly finds frequent item sets, then generates association rules from frequent itemsets (Yanna et al., 2011). The classical algorithms of association rules include Apriori algorithm and FP-Growth algorithm. For the Apriori algorithm, all frequent itemsets are generated by recursive methods. The pseudo code is as follows,

\[ L_1 = \{l | \text{arg}el - \text{itemsets}\} \]  
\[ For \left( k = 2; L_{k-1} \neq \emptyset; k++ \right) \]  
\[ C_k = \text{Apriori - gen}(L_{k-1}) \]
Start with all speed execution

\[ C_1 = \text{subset}(C_i,t) \]  

(4)

Execute \( c.count + + \) on all \( C_i \)

\[ L_k = \{ c \in C_k \mid c.count > \text{minsup} \} \]  

(5)

The Apriori algorithm needs to generate a large number of candidate frequent sets before generating frequent pattern sets, and scans the database many times which makes the Apriori algorithm more complex in time and space (Krishna et al., 2014). In contrast, the FP-growth algorithm does not require frequent connecting with database in the algorithm of scanning the database only two times to process, which greatly reduces the time of scanning the database and the database connection and the time and reduces the complexity of the algorithm, solution cost and saves a lot of time.

2.2 Cluster analysis

Clustering and classification are different, in which the classification rules of clustering depend on the data itself, and the classification rules are predefined. However, the cluster analysis requires the variance of the data set itself. In fact, the distribution of data sets is different (Hong et al., 2009). If the distribution of the data set itself is not related to the prior assumptions, the result of the algorithm will be meaningless. Therefore, in the face of practical problems, it is an important step to analyze the clustering algorithm which is suitable for the study of this topic. Method of grid based clustering and the global space quantization set into a finite element vector space unit into a frame to form a grid structure; and then use the grid structure to complete the clustering, clustering operation all in the quantized space (Du et al., 2008). The projection data of high dimensional space has a low dimensional the processing time and space, and only the quantization space elements in each dimension are related, so the algorithm is fast. Typical methods based on grid clustering include CLIQUE and Wave-Cluster. K-Means algorithm is a widely used clustering algorithm, which belongs to the partition method as shown in Figure 3.

![Figure 3. Details of K-Means algorithm](image)

2.3 Classification analysis
Classification analysis is through the analysis of the training data set from the input by the classification algorithm, which is also called the sample database or the training set to "learn" to construct the classifier. And then we establish the model for each category classification analysis, classification analysis model and then trained on the new input data to classify new data (Thuraisingham et al., 2012). The input set of classification analysis is a set of records to be classified and several classes of feature vectors. A record of training concentration is called a sample. Data classification is divided into two steps: learning and training, as shown in Figure 4. Decision tree, just as its name implies, its data structure is definitely tree structure, its basic composition domain flow chart is similar. From the root of the tree down, each edge represents a two yuan to make decisions, after several judgment, form a tree structure, a decision tree node, branches and leaves, nodes and branches that similar rules to determine under what conditions will be like what value.

![Diagram of classification analysis](image)

**Figure 4.** Details of classification analysis

### 3. EXPERIMENT ANALYSIS

![Graph of data mining complexity vs System effectiveness](image)

**Figure 5.** Data mining complexity vs System effectiveness

The recommendation algorithm based on association rules uses the historical data of other users to recommend books. In this system, this module can display detailed information on the page of the book, a column called
"love XXX people love the area will recommend users associated with the other books of the book list, such as search" in the book search Java programming ideas ", and then click the book into the detail page Java programming" book, on the page in addition to detailed information, score, book readers are also evaluated, to provide a display area for other users may be interested in the book. An important part of the book also expresses the personalized recommendation module, which is an important supplement to the collaborative filtering algorithm. Because the new purchasing books lack of lending records, the system uses items to calculate the similarity (Zheng, 2014). And each new recommended value and new book top K will purchase new books in the library are recommended on the home. The system displays the final search results. Pagination displays the matching keywords in the page showing the results, and the user can manually adjust the paging size, sorting and sorting options. In the process of retrieval of books, the results will be grouped according to the summary of publishers and authors, and results in the left side of the page display, books and the number of each author items each. For each item of the book display, it will display the attribute information of the book, such as press, publication date, author and other information.

It can be seen from Figure 5 that the efficiency of the system platform using data mining will be greatly improved with the complexity of data mining algorithms. Compared with the method proposed in this paper, the system efficiency of the traditional method under the same algorithm complexity is lower than the data mining algorithm proposed in this paper.

![Figure 5](image_url)

Figure 5. Efficiency of data mining platform

It can be seen from Figure 6 that the cost of the system platform based on data mining will increase gradually with the increase of system users. And it can be seen that the cost of the system will increase slower and slower with the increase of users. Therefore, the cost of the system will gradually reduce with the increase of the degree of promotion, thus it is conducive to the widespread promotion of the system.

4. CONCLUSIONS

In this paper, we introduce the algorithms of classification, clustering, recommendation and so on, and then compare and analyze the performance and the scope of the algorithms. Finally, we choose the K-means clustering algorithm and Apriori's frequent item discovery algorithm and apply them to the system. The readers clustering, book clustering, readers borrowing behavior analysis and other system functions are introduced. User-centric system is designed to distribute part of the administrator rights to individual users, so that users can also participate in the management of activities under the permission to allow both administrator time and cost savings, but also allow users to get a better sense of experience.
REFERENCES


Qinghua J. (2016). Data Mining and Management System Design and Application for College Student Mental Health, Intelligent Transportation, Big Data & Smart City (ICITBS), 2016 International Conference on. IEEE, 2016, 410-413.


