

Data mining Technology-Based Design and Implementation of University Library Personalized Service

Ye Lipeng^{1, a}, Yongfu Wang^{2, b}

¹Art Design Department of Wuhan Technology University

²Wuhan University of Technology Library

^a19671582@qq.coml, ^b19671582@qq.com

Keywords: data mining; personalized service; cluster analysis; association rules

Abstract. With IT and network technology becoming more and more mature, the application of database technology is deeper, university library data storage and management also significantly increase during this process, how to expand the space and scope of university library services so that readers can access information resources they need is put on the agenda, university library of personalized service is born. In order to make university library personalized service able to have a strong theoretical work as a basis, the paper bases on data mining technology, focuses on cluster analysis and association rules models, analyzes and researches the library personalized service.

Introduction

In today's era of the rapid development of information and digital technology, the management of university libraries data is increasingly complex and diverse, in order to meet the needs of the times, management of university libraries also tends to digital and information, while the data mining technology is precisely to solve this problem, providing a good platform for readers to quickly search out the needed date in such a large amount of library data. Through cluster analysis and association rules model, data mining technology transforms the traditional library services to personalized and humanized services, to facilitate the reader to retrieve and read.

Cluster analysis

According to some rule, divide a large amount of clutter data stored in database into a plurality of data sets, the data in each data set are not the same, but there is the similarity of the date in the same data set. Such clustering analysis on the establishment of a macro concept, readers in search of information you can consult the relevant literature to more relevant content, convenience of the reader effectively summarized.

The association rule model

The value associated with the name suggests is there are some laws between two or more variables on the known presence of these variables are associated. Association rules between the large amount of data is from the collection of items to find meaningful association, thus achieving the objective laws of technical methods of understanding of things.

Definition: Set $I = \{i_1, i_2 \dots, i_m\}$ is the set of all projects, set A is a collection composed by the project, called item-sets. Transaction T is a subset of items, each transaction has a unique transaction identifier Tid . Transaction T contains item-set A , if and only if AT . If the item set A contains k items, claimed it as k item set. D is a transaction database, the number of items appearing in the set A transaction database D , D as a percentage of the total transaction is called term support sets (support). If the support of the item set exceeds a user-specified minimum support threshold, the set is said to frequent item sets (or large items).

Support refers to the frequency pattern appears in the rule, if the transaction database has $s\%$ of transactions contain XY , XY called association rules in D , the degree of support for the $s\%$, in fact, can be expressed as the probability $P(XY)$, That support $(XY) = P(XY)$. Trust refers to the

inherent strength that matters $c\%$ of transactions in D containing X contains XY . If the support of X is support (x), as is the trust rules: support (XY) / support (X), which is a conditional probability $P(Y | X)$, i.e. confidence (XY) = $P(Y | X)$.

Traditional association rules algorithm

Apriori algorithm is an association rule mining frequent item-sets algorithm, the core idea is to dig down through the closed frequent item-sets candidate sets generated in two phases to detect and plot. And algorithms have been widely applied to business, network security and other fields.

Optimized Apriori algorithm

By Apriori algorithm analysis can be seen, how to reduce the number of visits to the transaction database and reduce the generation of candidate sets is crucial Apriori algorithm efficiency. In this paper, Apriori algorithm is improved, the main use of the hash table and compression technology portfolio item sets, thus improving the efficiency of the system. After construct a hash table count in the calculation of support, you only need to scan a portion HASH table rows do not need to scan all all of SASH table, thereby improving the efficiency of Apriori algorithm to achieve efficient mining frequent item-sets purposes.

The design of personalized system architecture model

In the paper, the overall design model of university library personalized service system is shown in Figure 1, the structure chart displays the user modeling according to the overall users information attributes from the collection of user information to the use of relevant technology, and finally matches the user model to the rule base which has been generated , then provides personalized information services to users.

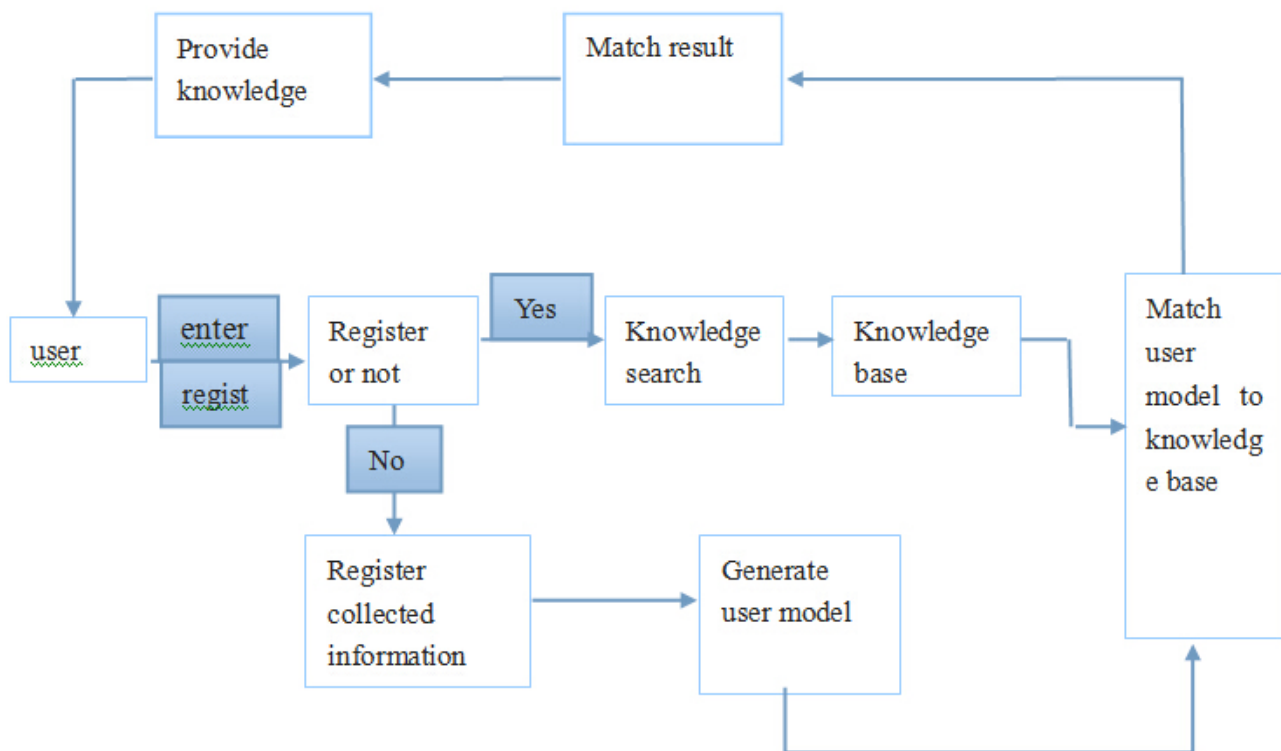


Figure 1 Structure chart of data mining-based university library personalized service system

After the user Sign libraries personalized service system, according to the system prompt, if already a registered user, you can directly enter the required knowledge retrieval based on personal knowledge, if the search results coincide with the user model, the system returns to the user personalization match results, and to provide the appropriate knowledge of the user. If unregistered users will need to follow the system prompts to register a personalized collection of information,

and then generate the user model, and then the user model and the knowledge base to match, if the two coincide, the relevant data prompt is returned.

Realization of data mining technology in the university library personalized system

According to the above model design, this chapter practices data mining techniques theory of university library personalized system in reality, the operation flow is shown as chart 2 (data mining process). The computer records and keeps user's basic information, loan history, search history records and bibliographic information, then the system pre-process and converse these data, store to the database unitedly, then the system through data mining technology for user information, loan historical information, historical and bibliographic information retrieval mining research to identify the hidden information between association rules, and finally this association is reflected in the user interface aspects of the user search and find, improve operational efficiency Library. Among them, the main law at the time of the database information mining to follow is this: on the one hand through data mining to maximize the optimization of the library's books, magazines and other resources, including books placed behind the museum library type, number and so on. Base on the effort to meet the needs of users, save library space and textbook funds to prevent waste books. On the other hand is personal expression, making users find the books most needed in the shortest possible time.

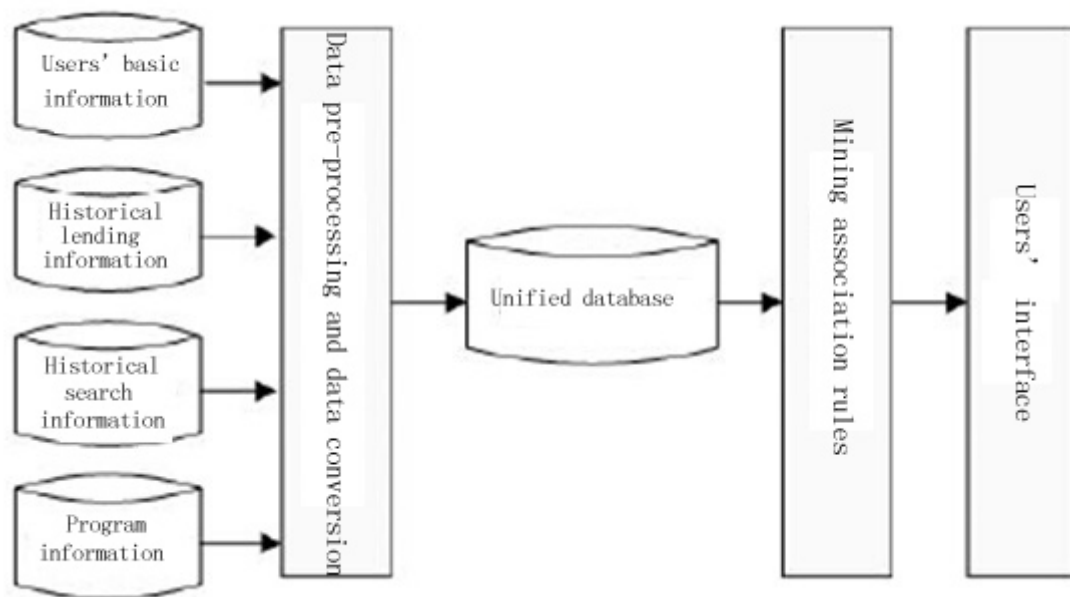


Figure 2 data mining process

The function description of the university library personalization system

The basic functions of the system mainly include three aspects, namely: First, collect behavior recording module of user information, mainly for library information collection and readers information gathering; second, model analysis of reader retrieval and borrow, through a comprehensive analysis of collected data, identify readers' borrowing law, building readers' borrowing model; and third, recommend retrieval and borrowing module, which is the ultimate module all the researches want to achieve, an application module of practical significance, the paper bases on association rules model to conduct the build of recommended module.

Conclusion

Through the research of data mining technology and the application of clustering algorithm and

association rules algorithm, we found a regular pattern, providing readers with personalized services. However, despite our significant progress in terms of theory and experiment of data mining technology-based university library personalized system, we are still limited by the development of computer technology and network technology, our personalized university library system still exists many deficiencies in the actual operation, the achievements of this paper is only a preliminary study of the library personalized service, many problems still exist, so we still have to carry out in-depth research on the personalized service of library.

References

- [1] Wang Fang. Data Mining Technology-Based Digital Library Personalized Service [J]. Heilongjiang College of Education, 2013,05: 185-186.
- [2] Liang Keyuan. Data Mining-Based Personalized Service Research of University Library [J]. Inner Mongolia Science Technology and Economy, 2008,17: 200-201 + 203.
- [3] Wei Ping. Explore Application of Data Mining-Techniques in the University Library Personalized Service [J]. Inner Mongolia Science Technology and Economy, 2014,15: 141-142.
- [4] Zhang Yanfang. Data Mining Technology-Based Library Discussion of Personalized Information Services [J]. Agricultural Library and Information Sciences, 2010,11: 258-260 + 264.
- [5] Nie Feixia. Data Mining Technology-Based Mobile Library Personalized Books Recommendation Service [J]. Library Science, 2014,05: 104-106.