Design and Implementation of k-most Library Users Mining

Shaik Mahaboob Basha[1], Pilli Dharmendra Kumar[2]

[1]Computer Science Engineering, Rise Group of Institutions, Ongole, India

Abstract: *Since library storage has been increasing day by day, it is difficult for readers to find the books which interest them as well as representative booklists. How to utilize meaningful information effectively to improve the service quality of the digital library appears to be very important. The process considered not only the relationship of a cluster of users but also the associations among the information accessed. This paper not only constructs a recommendation service for readers to search books from the web but takes the initiative in finding the most suitable books for readers as well. Furthermore, library managers are expected to purchase core and hot books from a limited budget to maintain and satisfy the requirements of readers along with promoting digital library services. The experimental results of association rule mining in academic libraries revealed interesting rules with respect to users interest in usage of books. The rules also help the institutions for arriving decision related to investment of budget for specific books titles.*

Keywords: data mining, knowledge management, association rule mining, library circulation data, hidden patterns, library

1. Introduction

The digital revolution has made digitized information easy to capture, process, store, distribute and transmit with significant progress in computing and related technologies and their ever expanding usage in different walks of life, huge amount of data of diverse characteristics continue to be collected and stored in databases. The rate at which such data are stored is growing phenomenally, where large volumes of data are stored in centralized or distributed databases. An academic library is an organized collection of user transactions stored in large databases [2, 4]. Discovering knowledge in large databases is well-defined process consisting of several distinct steps. Data mining is the core step, which results in the discovery of hidden but useful knowledge from massive databases. A formal definition of Knowledge discovery in databases is given as follows: "Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data" [3].

Data mining techniques when applied to library transactional databases helps to explore interesting patterns of data that help the librarians to take decisions on usage of books patterns, the kind of books and magazines that has to be purchased. The user accessing patterns when analyzed help to provide better services such as increasing library timings, providing them with additional resources for academic references. The data mining techniques in library also helps to identify the hidden knowledge that is stored in the databases, which helps the librarians and organization to take decisions based on their own data for providing better services and effective utilization of resources. The objective of this work is to analyse the user access patterns in library book circulation data using association rule mining task.

The paper is organized as follows. Section 2 describes knowledge discovery process in data mining and also describes the data mining tasks used for KDD analysis. In Section 3 an overview of the current library scenario and the need for data mining technique is discussed. Section 4 presents with methodology applied for mining association rules. Section 5 discusses the experimental results and the conclusion in section 6.

2. Knowledge Discovery In Database Introduction

Knowledge discovery in databases is defined as the nontrivial process of identifying of valid, novel, potentially usefully and ultimately understandable patterns in data.
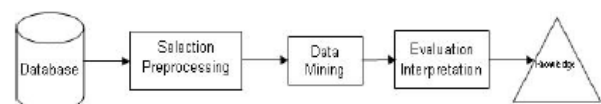

Figure 1: KDD Process

KDD process contains the following steps:

- Understanding the application domain: This includes relevant prior knowledge and goals of the application for which data mining tasks are applied.
- Extracting the target dataset: Selecting a dataset or focusing on a subset of variables using selection techniques.
- Data preprocessing: This is required to improve the quality of the actual data for mining. This also increases the mining efficiency by reducing the time for mining the preprocessed data. Data preprocessing involves data
- cleaning, data transformation, data integration, data reduction or data compression for compact representation etc.
- Data Mining: Data mining tasks such as association rule mining, clustering, classification are applied to find the hidden knowledge.
- Data Evaluation: This stage is used to analyse and evaluate the outcome of the data mining tasks for inferring knowledge.

Paper ID: SUB159029

146

3. Data Mining For Library

The current libraries make use of the information technology effectively for providing easier management of data and retrieval of information. Most of the libraries data is maintained in the databases, where day to day functioning of library operation like issue, retrieval, searching, querying are carried out. The outcome of the information stored in the database by the libraries are not audited are gone through for any purpose. The current information retrieval systems used in libraries help to carry out the management of work easier, which was manual earlier. The data in the current system helps to query the information which has to be known by the user stating what has to be put it in to the system to get an answer. The limitation of the current system in inferring knowledge is as given below

- An user has to post a query to retrieve a book either by specifying the title, author name, publication etc, in which he has to be familiar with any one of the key term.
- The librarian when need to identify the issues of a particular book frequency he has to manually provide the data for searching and processing, where the system has the data related to it.
- The knowledge about the users interest related to particular titles cannot be queried because the information is not directly stored but hidden somewhere.
- The libraries spend lakhs of amount in investing in resources and magazines and journals. The out come of the investment in terms of cost benefit analysis cannot be predicted.
- Identify the services that can be provided by libraries to cater the user needs in academic and research is not possible.

To infer the knowledge from the databases of the library system data mining plays a prominent role to predict patterns and infer rules, which helps for inferring decision through the knowledge gained are listed below.

- Increase books borrowing count for students.
- Understand user behaviour and patterns of book access
- Identify the resources of large number of users interested in.
- Make investment accordingly on the identified areas.
- Cut down cost on unused magazines and journals.
- Provide better services to help the users in academic and research.

4. Methodology –Mining User Interest Using Arm

This section describes the methodology used for mining users interest in library circulation data using association rule mining technique. The flow of methodology for the work is given in Fig. 2. Methodology for mining user patterns consists of important stages such as Preprocessing, Association rule mining and knowledge discovery.

The library circulation data for books transaction is used as dataset. The dataset consists of 1 lakh instances for one year data with 45 attributes.
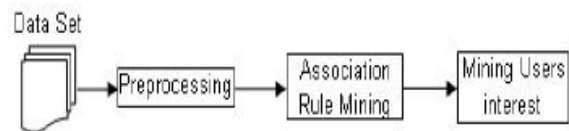

Figure 2: KDD Process

4.1 Preprocessing

In the preprocessing phase the circulation data retrieved from the datasets book_details, user_details, book_transaction are integrated with necessary attributes for association rule mining phase. The attributes like edition, publisher, price, issue_data, return_data, fine-amount, user-name etc., which are not relevant for this study are omitted.

4.2. Association Rule Mining (ARM)

Association rules are implications of the form X -> Y where X and Y are two disjoint subsets of all available items. X is called the antecedent or LHS (left hand side) and Y is called the consequent or RHS (right hand side). The association rules are used to predict correlation between the data items [1, 2]. The preliminary definitions and the metrics of association rules mining like Support, Confidence, Lift, Leverage and Conviction are presented below.

Confidence is the proportion of the examples covered by the premise that are also covered by the consequence (Class association rules can only be mined using confidence). Lift is confidence divided by the proportion of all examples that are covered by the consequence. This is a measure of the importance of the association that is independent of support. Leverage is the proportion of additional examples covered by both the premise and consequence above those expected if the premise and consequence were independent of each other. The total number of examples that this represents is presented in brackets following the leverage. Conviction is another measure of departure from independence. Conviction is given by P(premise)P(!consequence) / P(premise, !consequence) [5].

The association rules are inferred to predict patterns for inferring decision related book purchase and budget allocation with respect to specific titles.

5. Conclusion

Libraries have to satisfy user's information needs for knowledge sharing. The advancements in information and communication technology has provided with lot of resources, where people coming to the libraries are reduced. To attraction the users the libraries are expected to understand the needs of them and have to provide the services accordingly. The automation in library system used in libraries is very efficient in maintaining the library by reducing the manual work where there exist no methods to understand the data pattern stored for a long period of days. From the experimental results it is found that data mining techniques helps to identify the patterns, by bringing out the hidden knowledge stored in the databases for the effective and efficient use of libraries. The conclusion is to state that

Paper ID: SUB159029

147

the data mining is an important need for libraries to bring out the hidden knowledge.

References

[1] R. Arrawal, and R. Srikant, "Fast algorithms for mining association rules in large database," in Proceeding of the 20th International conference on very large database, 1994.

[2] V. Bhuvaneswari, and C.Esther Tamilselvi, "Students Placement Competency Analysis – An Apriori Framework", Journal of CiiT International Journal of Data Mining and Knowledge Engineering, September 2009. DOI: DMKE102009005

[3] Chuen-Min Huang, Sheng-Hsiu Kang, Ching-Che Chang, and Sheng- Hsing Lu. (2007). Apply Data Mining Techniques to Library Circulation Records and Usage Patterns Analysis. [Online] Available: http://140.125.84.58:8080/TeacherWeb/fileDownload/55.pdf

[4] S.-C. Kao, H.-C. Chang, and C.-H. Lin, "Decision support for the academic library acquisition budget allocation via circulation database mining," Information Processing and Management, vol. 39, pp. 133– 147, 2003.

[5] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten, The WEKA Data Mining Software: An Update; SIGKDD Explorations, vol. 11(1), 2009.

[6] R. Porkodi, V. Bhuvaneswari, R. Rajesh, and T. Amudha, "An Improved Association Rule Mining Technique for XML Data using XQUERY and Apriori Algorithm", in Proceeding of the conference on IEEE International Advance Computing Conference, pp. 1510-1514, 2009.