

Frequent Pattern Mining Technique for Improving Book Lending Recommendation Service

Swasti Saxena¹, Bhawana Singh², Zubair Khan³

^{1,2}M.Tech 2nd, Invertis University, Bareilly (U.P)

³H.O.D (C.S), Invertis University, Bareilly (U.P)

Abstract: In data mining Association rule is an important technology. It helps in finding associations, frequent patterns, correlations or causal structures among set of items or objects in transactional database, relational database and other information repositories. In library, this technique helps in recommendation of best books related to a particular field to the students. It helps in decision support for the library services.

We do here the case analysis of library services and proposed an idea to solve the problem of students who have confusion in issuing a book related to a particular field. As there are a number of books related to a particular field, so, students get confused that which book will be better and which book he or she should issue. This problem is solved here by using a technique which is known as Enhanced Frequent Pattern Technique (EFPT). By this technique, students will get to know which book he should issue among a number of books related to a particular field.

Keywords: data mining, library users, EFPT, resource recommendation.

1. INTRODUCTION

With the excessive use of database system, a huge historical data is accumulated in university library. There exists important information (hidden knowledge) behind these data. We hope that some methods can be used to analyze data up to high level, so useful information can be got and applied to user personal service. How to mine and utilize the user behavior characteristic in order to further enhance the library services level has become a new problem.

Data mining is a term used in computer science. It helps in finding new information in a large amount of data. Data mining is also known as knowledge discovery in databases (KDD) process. The technology of storage and data collection has made it possible for organizations to accumulate large amount of data at lower cost.

The system will produce large amounts of data every day, but these data are usually used only for general traffic statistics lacking of integration and analysis. Therefore, people are eager to have a technology which can provide us with

automatic and intelligent search, and the data mining just meets the requirements in this regard. In the five functions of the data mining, the association rule analysis can be a good method to unearth a lot of data and discover rules and characteristics of readers, to provide guidance for carrying out the library service.

In data mining, association rule plays a major role as it can be used in various fields according to the requirements. It can be used in library services to trap the borrowing information of the book borrowers. With the help of this technique, the efficiency of library management can be increased.

Market Basket Analysis- This process analyzes the customer buying habits by finding associations between the different items that customer place in their “shopping baskets”.

There are two major considerations of apriori algorithm:

1. Support and
2. Confidence.

2. DATA MINING

Data mining is a term used in computer science. It helps in finding new information in a large amount of data. Data mining is also known as knowledge discovery in databases (KDD) process. The technology of storage and data collection has made it possible for organizations to accumulate large amount of data at lower cost.

Types of Data to be mined- Data mining is not bounded to one type of data. It should be applicable to any kind of data. The different approaches and algorithms are applied depending on the kind of data. There are various types of data in which mining is performed. They are:

- Relational databases
- Flat files
- Data ware houses

- Transactional databases
- Multimedia databases
- Spatial databases
- Time series databases and
- Web mining.

The process of exploration and analysis by auto or semiautomatic means and discovering meaningful patterns and rules is known as data mining. It involve a computational process which helps in computing large data set's pattern discovery. The main goal of this advanced feature is to extract information from a data set and transform it into desirable structure for further use. KDD is an extraction of previously unknown and potentially useful information from data in databases. Though data mining and KDD are sounds as same but data mining is actually a part of the knowledge discovery KDD.

Data mining functionalities- the data mining functionalities can be categorized as:

- Characterization
- Discrimination
- Association rules
- Classification
- Prediction
- Clustering
- Outlier analysis and
- Evolution and derivation analysis.

There are two main purposes of data mining-

1. To create a descriptive model
2. To create a predictive model.

A descriptive model shows the main characteristics of the data set. It is the summary of data points making it possible to study the important aspects of data set. A descriptive model is found through undirected data mining. It is also known as bottom up approach.

The predictive model is used to predict an unknown value of a specific variable. The predictive model is created from given known values of variables.

3. LIBRARY USERS

There are a number of types of users who make use of library particularly according to their needs. As we consider a college having a number of courses running in it like B.Tech, M.Tech, BBA, MBA, PGDM, etc. All the students search book

differently in the library. Some students make a search of books by their authors name, some search books by the book title name or some search by the publication of books. So in library management there should be the consideration of all above things so that student gets help from this.

4. EFPT

4.1 Overview of FP-Growth

Fp tree compresses a large database into a compact tree structure. It avoids in repetitive database scan. It helps in developing an efficient; fp tree based frequent pattern mining methods. It can be said a type of divide and conquer methodology in which it decompose mining tasks into smaller ones. This technique helps in avoiding candidate generation. It do only sub database tests.

FP-tree is a **frequent pattern tree**. Formally, FP-tree is a tree structure defined below:

1. One root labeled as "null", a set of *item prefix sub-trees* as the children of the root, and a *frequent-item header table*.
2. Each node in the *item prefix sub-trees* has three fields:
 - Item-name: register name of the node represents.
 - Count: the number of nodes lying upto the destination node.
 - Node-link: links between the nodes.
3. Each entry in the *frequent-item header table* has two fields,
 - Head of node-link and item name which points to the first node in the FP-tree related to item name.

Compress a large database into a compact, *Frequent-Pattern tree* (FP-tree) structure

- highly compacted, but complete for frequent pattern mining
- avoid costly repeated database scans

Develop an efficient, FP-tree-based frequent pattern mining method (FP-growth)

- it uses divide-and-conquer methodology: decompose mining tasks into smaller ones

it avoid candidate generation test: sub-database test only.

Construct FP-tree

Two Steps:

1. Scan the transaction DB for the first time, find frequent items (single item patterns) and order them into a list L in frequency descending order. e.g., L= {f: 4, c: 4, a: 3, b: 3, m: 3, p: 3}. In the format of (item-name, support)

- For each transaction, order its frequent items according to the order in L; Scan DB the second time, construct FP-tree by putting each frequency ordered transaction onto it.

5. BOOK RECOMMENDATION USING EFPT

As it is well known that library is a collection of different books of different areas. There are a number of books of a particular subject with different authors, so students get confused in the different books related to a particular field.

With the help of frequency of issued books, we suggest a technique which helps him in this problem.

Table 1. Storage of number of items

Book Title Name	Number of Books
Data mining	45
Computer networks	40
Distributed systems	33
Theory of automata	38
Software engg.	48

Let us suppose a student want a book of data mining and he has confusion between two Authors. He wants to know which book is best suited to students. Then here in this paper he can get knowledge about this. He just enters both the authors of data mining book in which he has confusion one by one.

By this he gets the details of both the book that is how many times a particular book of his choice has been issued. Now he can compare and get a clear idea in issuing any book. As it is clear, the book issued many number of times should be better than that less number of times.

BOOK NAME-DATA MINING

AUTHORS-KAMBER & STEINBACH

Table 2: Database of Books

Author Name	Book_id	Book Title	Frequency
Micheline kamber & jian Pei	dm1111	data mining concept & techniques	17
Arun k. pujari	dm1112	data mining techni.	8
Steinbach	dm1113	intro. To data mining	6
Chattam velli	dm1114	data mining algo	3

Fig. 1. Database of Books

When student enters author name=KAMBER

Table 3. Searching with author name

Author Name	book_id	Title of book	frequency
Kamber	dm1111	data mining concept & techniques	17

When student enters author name=STEINBACH

Author name	book_id	title of book	Frequency
Steinbach	dm1113	intro. To data mining	6

“Book Search” page having details of all books. To make the search faster select the “Search By” option in which four categories of search are given: Book Id, Book Name, Author and Publication.

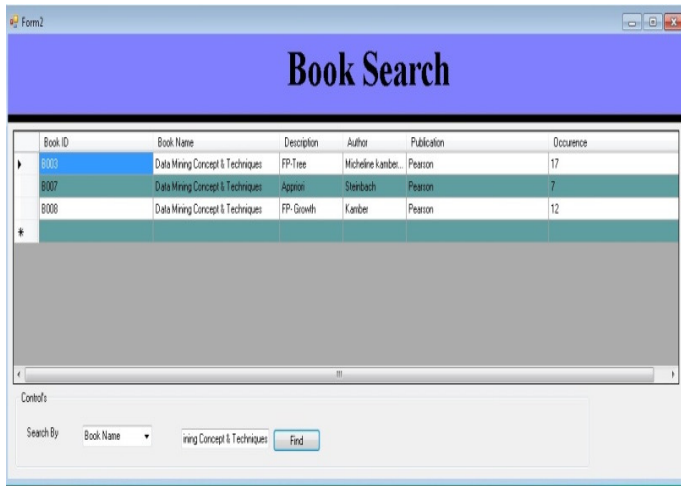


Fig. 2: Searching of Book

6. CONCLUSION

Concerning about our library borrowing issues, it is clear that the student who has confusion in two books related to a particular subject can get a better idea by this technique. The searching time of book is also less as compared to that of apriori algorithm and the result of finding the best book among a number of books related to a particular field is also done here in a different way with the help of frequent pattern tree. In our considered example of proposed methodology it is clear that if one book is issued 17 times and the other only 6 times the student will go with the first choice.

Thus by this technique it will be beneficial for the students to get the best knowledge about different books and it also reduces the searching time to get the best possible output.

Hence in this paper, we remove two problems of again and again generating candidate set and reduce the time of getting results so that searching becomes fast. To make frequent pattern mining an essential task in data mining, much research is needed to further develop pattern-based mining methods.

REFERENCES

- [1] Gong Lunfeng, Lei Huan, Zhu Li, "The Application of Association Rules of Data Mining in Book-lending Service", 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012).
- [2] JianWei Li, PingHua Chen, "The application of Association rule in Library system", IEEE 2008.
- [3] Peng Peng, Qianli Ma and Chaoxiong Li, "The Research and Implementation of Data Mining Component Library System", World Congress on Software Engineering IEEE 2009.
- [4] Dianjun Xu, Min Sui, "Research on Application of Digital Library based on Association Rules", IEEE 2011.
- [5] Han, J., Pei, J., Yin, Y.: Mining Frequent Patterns without Candidate Generation. ACM SIGMOD (2000).
- [6] Bao Chui-Mei, Wang Zun-Xin, and Bai Ru-Jiang, "Data mining technology and its application in the library", Journal of Intelligence, Xi An, 2004 (09), pp. 49- 51.
- [7] J.Han, M.Kamber. Data Mining Concepts and Techniques. China sMachine Press. 2005.
- [8] Shuzhi An, Database and Data Mining, Hsinghua University Press, Beijing, June 2005.
- [9] Pang-Ning Tan, Instruction to Data Mining. Posts & Telecom Press, May 2006.
- [10] Man Lou: "Study of Library Quality Management Systems", Southwestern Jiao Tong University Press, 2009 (in Chinese).
- [11] X. Zhu and I. Davidson: "Knowledge Discovery and Data Mining: Challenges and Realities", Hershey, New York, 2007.
- [12] M. Kantardzic: "Data Mining: Concepts, Models, Methods, and Algorithms", John Wiley & Sons, 2003.