FREQUENT PATTERN MINING FOR EFFICIENT LIBRARY MANAGEMENT

ANURADHA.T
Assoc.prof, ECM Department
KL University
Guntur, INDIA
atadiparty@yahoo.co.in

SATYATEJ.K
ECM Department
KL University
Guntur, INDIA
satyatej.koganti@gmail.com

SRI SAI KRISHNA.A
ECM Department
KL University
Guntur, INDIA
saikrishna.gjc@gmail.com

NAGA ANIL KUMAR.G
ECM Department
KL University
Guntur, INDIA
anil.naidu58@gmail.com

***Abstract*—Data mining is used to extract meaningful information and to develop significant relationships among variables stored in large data sets involving methods from statistics and artificial intelligence but also management. In this paper data mining technique named association technique is applied to analyze the patterns which depicts the relationship between more frequently taken books by the students. The student's record of library data was studied based on many factors such as their subject requirements, number of books issued, and duration for each book. It is recommended that all these correlated information should be conveyed to the library department such that it helps them to improve the issue performance by avoiding the delay, maintain more number of highly required books according to the current subjects of the different streams of students.**

***Keywords*—frequent pattern; association technique; minimal support.**

I. Introduction

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data in order to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining)[1].Library management concerns with the management of resources which basically includes books, manuscripts, journals etc. and providing effective and efficient services to its users. Since the manual management of books and other resources in a library and keeping track of every books of the library accessed by the user is tedious job and hence often technological support is expected [7]. Further, to keep track of the books issued and returned across the counters, journals, periodicals and manuscripts consulted by the users and so on needs additional book keeping on additional parameters and are generally not done in a typical library which operates manually. The situation becomes acute when the library does procurement of resources.

II. Data Mining

Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making [7].Data mining software allow the users to analyze data from different dimensions categorize it and a summarized the relationships, identified during the mining process [1].Different data mining techniques are used in various fields of life such as medicine, statistical analysis, engineering, education, banking, marketing, sale etc.,

A. Frequent Patterns

Frequent item sets play an essential role in many data mining tasks that try to find interesting patterns from databases, such as association rules, correlations, sequences, episodes, classifiers, clusters and many more of which the mining of association rules is one of the most popular problems [6].The original motivation for searching association rules came from the need to analyze so called supermarket transaction data, that is, to examine customer behavior in terms of the purchased products [5]. Association rules describe how often items are purchased together.

B. Associations in Data-Mining

In data mining, **association rule learning** is a popular and well researched method for discovering interesting relations between variables in large databases [5]. Piatetsky-Shapiro describes analyzing and presenting strong rules discovered in databases using different measures of interestingness [3]. Based on the concept of strong rules, many introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) systems in supermarkets.
For example, the rule **{Onions, Potatoes} => {Burger}** found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy hamburger meat[3]. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements. In addition to the above example from market basket analysis[2] association rules are employed today in many application areas including Web usage mining, intrusion detection and bioinformatics.

C. Data mining in Library Management System

Library provides an essential element for the betterment and progress of the student career. It enables the students of a college/university to get updated in the academic subjects. Mining in educational environment is called Educational Data Mining, concern with developing new methods to discover knowledge from educational databases like library, sports, health and management etc., in order to analyze student's trends and behaviors towards particular subjects when in-close with examinations [1]. With the information deep and enough on management in library maintenance system will provide the student's to achieve quality objectives in their academic curriculum, data mining methodology which was used in Library Management can help bridging this knowledge gaps in higher education system.

TABLE:1

Student. No	CRYPTOGRAPHY AND NETWORK SECURITY	ANSCII-'C'	DBMS	DATAMINING TECHNIQUES
1	1	1	0	1
2	0	1	1	0
3	0	0	1	1
4	1	0	0	1

Sample table showing some of the books issued to students

In the above table, when the book is taken by the student it is considered as '1' and if not it is '0'.Similarly about 2000 different student records for seven different books are considered.

III. Proposed Model

In a university library overall books issued to a student is determined by the automation records of the university library. The proposed model mainly deals with finding relationship between the most frequent books that are taken by the students for that semester based on the conditions of stipulated time period and maximum books issued. In general the books were issued for a period of 15 days and providing an option for students to renewal those books for later period of time. While at the same time issued books of a student can get returned

in any of these 15 days. A student can take a maximum of four different books within that period. The proposed model makes prediction about most chosen books by students based on class automation as well as system information from the department.

We have considered about 2000 records of students, and the books considered were the important books required by the students for that particular semester. Initially, we calculated the count of the various seven books which are taken individually. Then the books which were less than the minimum support key were eliminated from further proceedings. Count of the books were as follows:

1. Computer Networks : 1098
2. Compiler Design : 965
3. Data Mining : 1065
4. Operating Systems : 688
5. Web designing : 257
6. Java : 219
7. Software Engineering : 896

From the above data we can identify that the count of both Web designing and Java is less than the minimum support count of 300.So they are eliminated from further proceedings of frequent patterns.

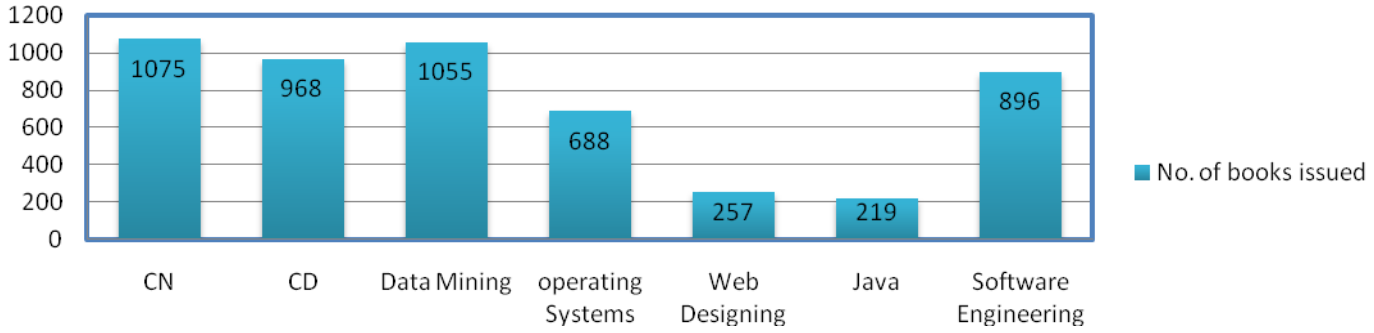

Fig.1. Representation of individual books issued to students

Now for the frequent patterns we have to consider the rest of books over which we are applying the data-mining techniques. On applying the frequent patterns we calculated what the count among 2000 students who have taken both the books together at the same time. After calculating the count of each frequent patterns, the later process was similar to the above process where the frequent 2 patterns are eliminated whose count was less than the minimal support key. Following data shows the count of various frequent 2 patterns:

1. COMPUTER NETWORKS and COMPILER DESIGN = 760
2. COMPUTER NETWORKS and DATA MINING = 870
3. COMPUTER NETWORKS and OPERATING SYSTEMS = 543
4. COMPUTER NETWORKS and SOFTWARE ENGINEERING = 396
5. COMPILER DESIGN and DATA MINING = 783
6. COMPILER DESIGN and OPERATING SYSTEMS = 287
7. COMPILER DESIGN and SOFTWARE ENGINEERING = 246
8. DATA MINING and OPERATING SYSTEMS = 340
9. DATA MINING and SOFTWARE ENGINEERING = 233
10. OPERATING SYSTEMS and SOFTWARE ENGINEERING = 269

Now the combinations COMPILER DESIGN and OPERATING SYSTEMS, COMPILER DESIGN and SOFTWARE ENGINEERING, DATA MINING and SOFTWARE ENGINEERING and OPERATING SYSTEMS and SOFTWARE ENGINEERING are not considered for further proceedings for less than minimum support.

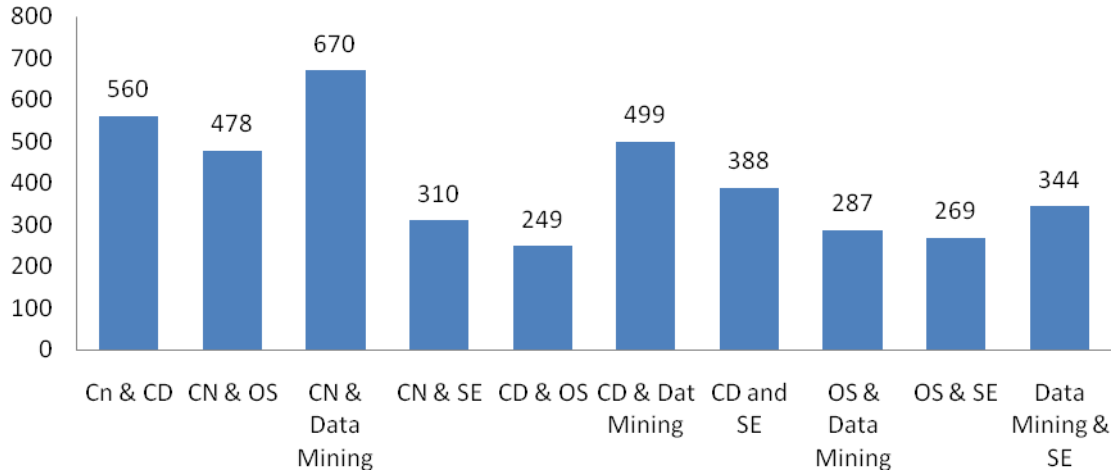

Fig.2. Representation of Frequent 2 pattern books

Finally, for the frequent 3 patterns we have only 2 different sets obtained (COMPUTER NETWORKS,COMPILER DESIGN,DATA MINING) and (COMPUTER NETWORKS,DATA MINING,OPERATING SYSTEMS) after considering the minimum support and minimum confidence values.

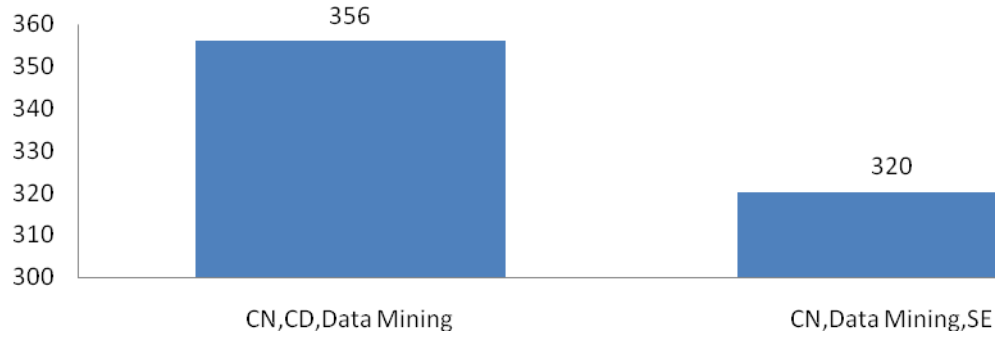

Fig.3. Representation of Frequent 3 pattern books

IV.Results

On applying the data mining techniques on different 2000 records available we came with frequent patterns between (CN,CD,DataMining) and (CN,DataMining,SE).We can conclude that among 2000 students 356 students take all the three books namely (CN,CD,DataMining) and 320 students take all (CN,Data Mining,SE). Further, frequent 4 pattern book set is not obtained as the count of the combinations is less than the required minimum support key value.

V.Conclusion

By this way of applying the association techniques over a library data, it will help the library management to buy and maintain those books that are much required and hold them in large number than those that are not much preferred by the students.So, this helps for maximum number of students to get those frequent pattern related books and also for efficient retrieval of books. User can retrieve the books which he needed in a faster way which can save a lot of time for him rather than just searching for various books.

VI.Future Work

We have applied the data-mining technique for proper functioning of Library management. This can be extended from 7 available books to many books and over many thousands of students' information so that library management of much required books and efficient retrieval can be done successfully. Similarly indeed having more books we can divide the books as per departments and can apply these techniques so that it would be beneficial for all the students of different departments.

References

[1] Anitha, V; Sunitha Jeevan, C; Kala, S. (2011): Library Management System Using Association Rule Mining. International Journal of Biotech Trends and Technology –may to june issue.

[2] Data on Market Basket analysis from wikipedia. http://en.wikipedia.org/wiki/Market_basket

[3] Fayyad, U.M., Piatetsky-Shapiro, G.;Smyth, P. From Data Mining To Knowledge Discovery: An Overview. In Advances In Knowledge Discovery And Data Mining, eds. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, AAAI Press/The MIT Press, Menlo Park, CA, 1996, pp. 1-34.

[4] Frawley, W.J; Piatetsky-Shapiro,G;Matheus,C.J. Knowledge Discovery in Databases: AAAI/MIT Press, 1991.

[5] Khattak,A.M;Khan,A.M;Sungyoung,L.(2010) : Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka. International Journal of Database Theory and Application Vol. 3, No. 1, March, 2010.

[6] R. Agrawal, T. Imielinski, and A. Swami, "Mining Associations between set of items in Massive Databases". In proc. Of the ACM-SIGMOD 1993 int'l conf. on Management of Data, pages 207-216, Washington D.C. USA, 1993.

[7] Shaeela,A;Tasleem,M;Ahsan,R.S;Inayat Khan,M. (2010): Data Mining Model for Higher Education System. European Journal of Scientific Research Vol.43 No.1 (2010), pp.24-29.

Authors Profile

[1] T.ANURADHA

Assoc.Prof and doing research in DATA BASE MANAGEMENT STUDIES in KL UNIVERSITY, GUNTUR

[2] A.SRI SAI KRISHNA

Student at KL University, studying IVth year B.tech in ECM Department

[3] K.SATYATEJ

Student at KL University, studying IVth year B.tech in ECM Department

[4] G.NAGA ANIL KUMAR

Student at KL University, studying IVth year B.tech in ECM Department