Discovery of Association Rules at Multiple Levels

Naveen
Department of Computer Science and Engineering
D A V College of Engineering and Technology, Kanina, Mohindergarh, Haryana (India)-123027 Hisar

Parveen Kumar
Department of Computer Science and Engineering
Guru Jambheshwar University of Science and Technology, Hisar, Haryana (India)-125001

Ruchika Yadav
Department of Computer Science and Engineering
Haryana College of Technology and Management, Kaithal, Haryana (India)-136027

Abstract - Data mining is extraction of implicit, previously unknown; potentially use for information from the vast amount of data available in the data sets (databases, data warehouses or other information repositories). In previous studies the association rules are generated at the single conceptual levels however mining association rules at multiple concept levels may lead to the discovery of more specific and concrete knowledge from large transaction databases by extension of some existing rules mining techniques. In multilevel association rules we use different minimum support for different conceptual levels. In this paper, multiple-level association rules are discussed using MLT2 algorithm. This algorithm discovers association rules for successive levels making use of rules already discovered for upper levels of concept hierarchy.

Key Words: Multiple-level Association rules, Data mining, Support, Confidence.

I. INTRODUCTION

Data mining refers to extracting or mining knowledge from large amounts of data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus, data mining should have been more appropriately named "knowledge mining from data," which is sometime known as knowledge mining. Many other term carry a similar or slightly different meaning to data mining, such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology and knowledge discovery from data or KDD [5]. The overall process of extraction useful knowledge from data is known as KDD. Data mining is a particular step in this process having specific algorithms for extracting the useful information from data [8]. The process of data mining is the use of algorithms to extract the useful information and patterns that the knowledge discovery process strives for.

Several Techniques for data mining are Association rules, Clustering, Rule induction, Classification, decision trees, Neural networks etc. Association rules in data mining are useful for discovering relation between items, Catalogue design, store layouts, product placement, target marketing etc. [5].

II. ASSOCIATION RULES

Association rules are heart of data mining. The association rules are detects hidden linkages of otherwise it is supposed unrelated data. These linkages are called rules [7].Association is the degree of relationship or involvement or the connection of objects. Such connection is termed as association rule. An association rules revels the associative relationship among objects at multiple levels.

III. APPLICATIONS OF ASSOCIATION RULES

I. Market Basket Analysis: Market Basket is mostly and typical example of association rule mining. Data collection using bar code scanner in market is an example of market basket analysis. In the 'market basket' database there are a large no of records. The data in the database is typically used for cross spellings, promotions, catalogue design and to identify customer [15].

II. Medical Diagnosis: Application rules are also applied by the physicians in medicines to cure the patients. In the medical diagnosis relational association rules are applied to identify the probability of illness and symptoms of new disease [13].

III. Census Data: census contains huge variety of statical information of society for both researchers and general public. The information related to population and economic services are forecasted in planning public services [12].

IV. CRM of credit card business: the preference of different customer groups, products and services are identified by banks using CRM. The collective application of association rule techniques reinforces the knowledge management process and allows marketing personnel to know their customers well to provide better quality services [11].

IV. MULTIPLE LEVEL ASSOCIATION RULES

In multiple-level association rule mining, the items in an item-set are characterized by using a concept hierarchy. Mining occurs at multiple levels in the hierarchy. At lowest levels, it might be that no rules may match the constraints. At highest levels, rules can be extremely general [3]. Generally, a top-down approach is used where the support threshold varies from level to level (support is reduced going from higher to lower levels) .Sometimes, at primitive data level, data does not show any significant pattern [6]. But there are useful information hiding behind. The goal of Multiple-Level Association Analysis is to find the hidden information in or between levels of abstraction. Two general requirements to do multiple-level association rule mining [1]:

1) Provide data at multiple levels of abstraction.

2) Find efficient methods for multiple-level rule mining

V. MLT2 ALGORITHM

Algorithm ML_T2L1: Find multiple-level large item sets for mining strong ML association rules in a transaction database.

Input: (1) T [1], a hierarchy-information-encoded and task-relevant set of transaction database, in the format of (TID, Itemset), in which each item in the itemset contains encoded concept hierarchy information, and (2) the minimum support threshold (min-sup[l]) for each concept level 1.

Output: Multiple-level large item sets.

Method: A top-down, progressively deepening process, which collects large item-sets at different concept, levels as follows. Starting at level 1, derive for each level 1, the large K-items sets, L [l, k], for each k, and the large item set, L[l] (for all k's), as follows [4] [10]: -

```
1)      for (l:=1;L[l,1] <> 0 and l < max_level; l++) do begin
2)  if l = 1 then begin
3)  L[l,1] := get_large_1_itemsets (T[1],l);
4)      T [2]:= get_filtered_transaction_table (T[1], L[1,1]);
5)  End
6)      else L[l,1] := get_large_1_itemsets (T[2], l);
7)      for (k := 2; L[l, k – 1] <> 0; k++) do begin
8)  Ck:=get_candidate_set(L[l,k–1)
9)  For each transaction t in T[2] do begin
10) Ct := get_subsets (Ck , t);  // Candidates container in t
11)  For each candidate c in Ct do c.support++;
12)  End
13)  L[l, k] := {c in Ck | c.support >=  minsup[l]}
14)    End
15)    L[l] := Uk L[l, k];
16)     End
```

VI. EXPERIMENTAL RESULTS AND ANALYSIS

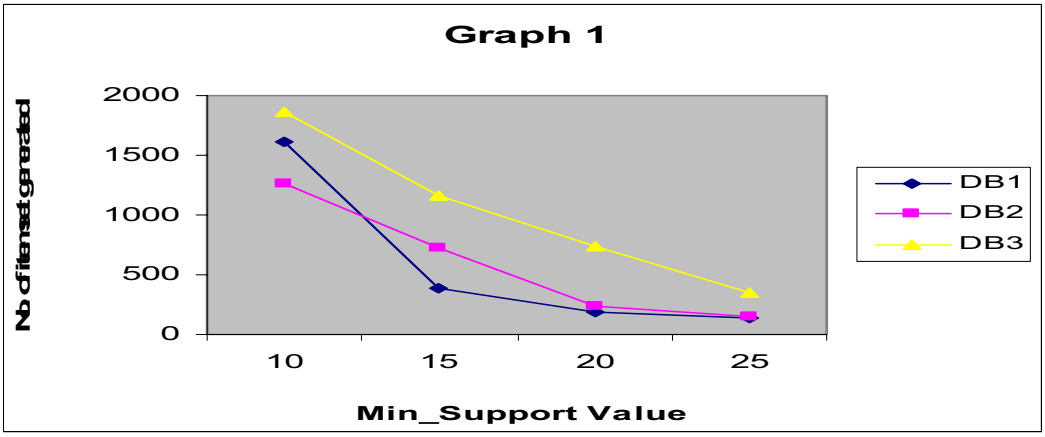
The data set named MeSH® used in this research work from real world domain. This datasets is available from NLM's PubMed database. The detail of this dataset is described below. [9]

S.No.	Dataset Name	Dataset Size	No. of Transactions	No. of Items
1.	MeSH-A	8MB	9,885	1800
2.	MeSH-C	10MB	10,000	830

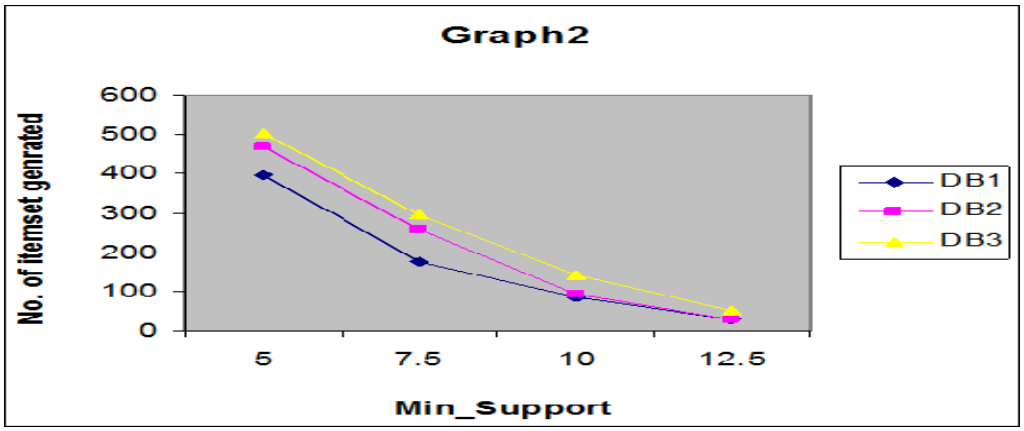
3.	MeSH-D	12MB	10,000	1200

Summary of Results using Min_Support and No. of Frequent Itemsets Generated Factor are given below:-
On the basis of Min_support and No. of frequent itemsets generated we have to draw the following graphs of the tables for analyzing the results. The graphs are as follows.

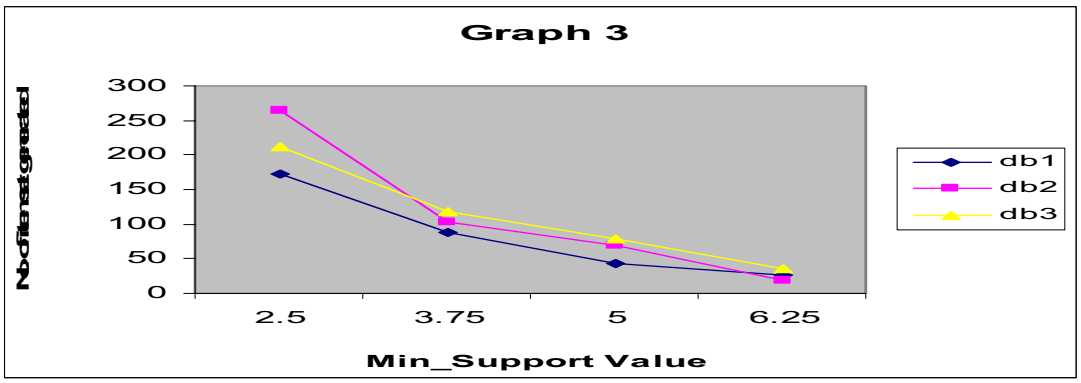
Parameters: Max_Support 90 %, Delta 0.5



Min_Sup Vs itemsets generated (Level-1)



Min_Sup Vs itemsets generated (Level-2)

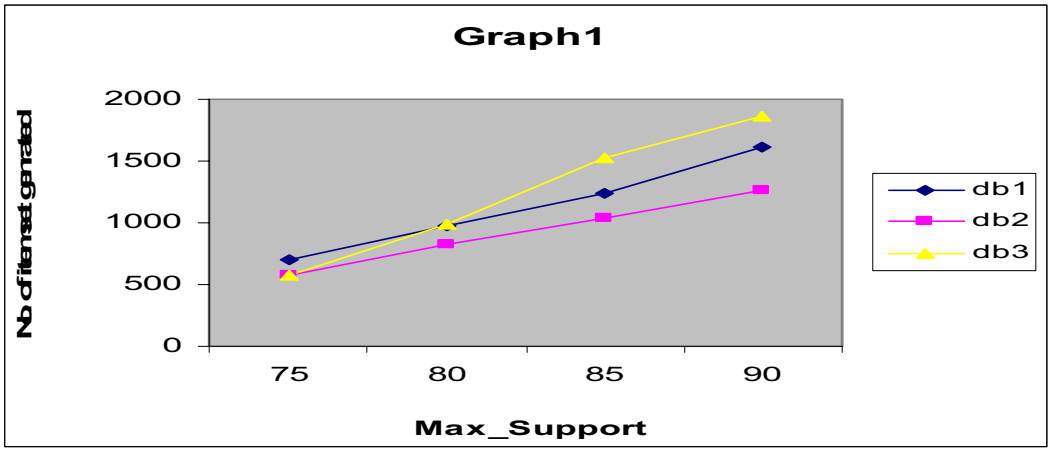


Min_Sup Vs itemsets generated (Level-3)

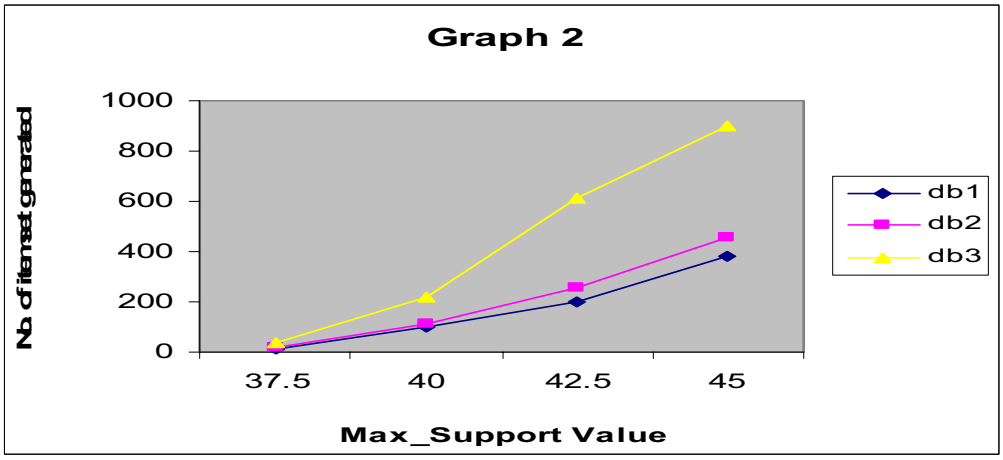
The above results shows that Min_Support is decreasing by factor delta(0.5).As Min_support decreases at lower levels we find very specific information. The execution time of the algorithm is variable for different datasets with a variation in Min_Support. The time for different frequent item set mining algorithms depends a lot on the structure of the data set. The mining of multiple-level rules can provide more specific information for the users at lower levels and enhance the flexibility and power of data mining systems.
Summary of Results using Max_Support and No. of Frequent Itemsets Generated Factors given below:-

On the basis of Max_support and No. of frequent itemsets generated we have to draw the following graphs of the tables for analyzing the results. The graphs are as follows.

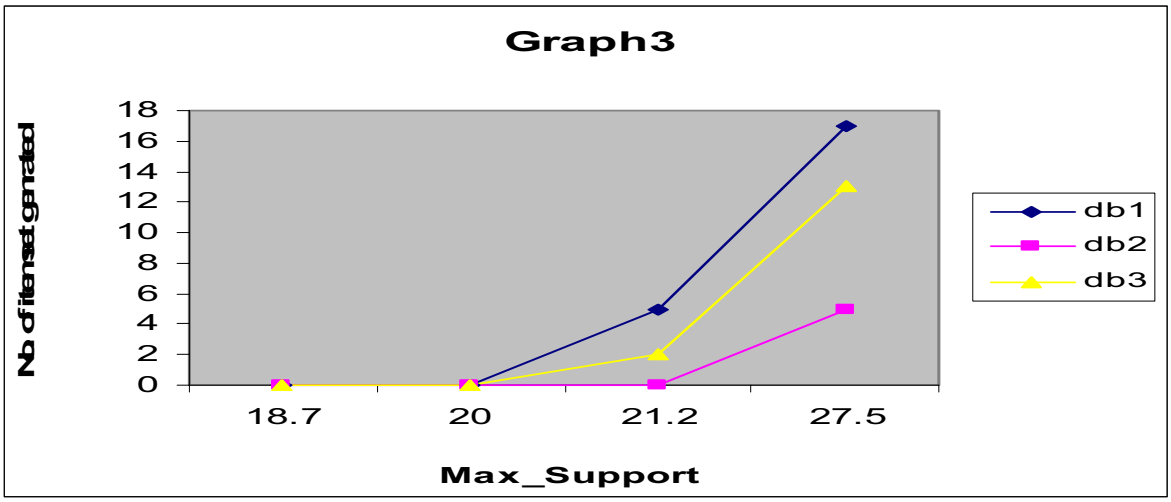
Parameters : Min_Support 10 %, Delta 0.5



Max_Sup Vs itemsets generated (Level-1)



Max_Sup Vs itemsets generated (Level-2)



Max_Sup Vs itemsets generated (Level-3)

The results shows that Max_Support is decreasing by factor delta (0.5). No. of frequent itemsets is decreases as Max_support decreases at each levels. Multiple-level association rules can be discovered efficiently from large databases.

VII. CONCLUSION

This study demonstrates that mining multiple-level knowledge is both practical and desirable. This work has successfully discovered multiple-level association rules using MLT2 algorithm. The association rules discovered abstraction. Our algorithm has efficiently discovered Multiple-level association rules from three datasets (MeSH-A, MeSH-C, MeSH-D) from NLM's PubMed database [9]. We have noticed that the execution time of the algorithm depends on the size and complexity of concept hierarchy discovered and hence it is variable for different datasets. This algorithm discovers association rules for successive levels making use of rules already discovered for upper levels of concept hierarchy. Number of association rules discovered depends on value of parameters at each level like support, confidence, and lift.

This work is contribution towards representing knowledge at multiple-levels in the form of association rules that enhances the ease and comprehensibility of the users.

VIII. FURTHER OPPORTUNITIES

In the field of association rule mining, most of the proposed methods for generating frequent patterns use the Apriori algorithm.Two main disadvantages to the Apriori approach are: First, the method may need to generate a large number of candidate sets. Second, repeated scans of the dataset to match and tally the patterns of candidates can be potentially time consuming where the dataset is large and/or the itemset is large [2]. So the FP-growth implementation would certainly be useful in such a situation [4].

REFERENCES

[1] A. Siddiqui, D. Prohaska A. Ghosh S. Anamanamuri, "Implementation of Multiple-Level Association Rule Mining in Weka ," June 2005.

[2] Janas and J.M., "A Priori Algorithm for Mining Multidimensional Association Rules," in Proc. 25th International Conference on Information Technology Interfaces IT1, June 2003, pp. 193-198.

[3] Choonho Kim and Juntae Kim, "A Recommendation Algorithm Using Multi-Level Association Rules," in Proc. WIC International Conference on Web Intelligence, Oct 2003, pp. 524-527.

[4] N.Rajkumar, M.R.Karthik and S.N.Sivanandam, "Fast Algorithm for Mining Multilevel Association Rules," 2003 IEEE, pp. 152-155.

[5] Jiawei Han and Micheline Kamber, "Data Mining: Concept and Tech-niques," 2nd Edition, 2001, pp.6-9.

[6] Jiawei Han and Yongjian Fu, "Discovery of Multiple-Level Association Rules from Large Databases," in Proc. 21st VLDB Conference, 1995.

[7] Lam N.S. "Discovering Association Rules in Data Mining" Department of Computer Science, University of Illinois at Urbana-Champaign [Online]. Available: www.raymond-lam.com/CS411.doc

[8] Usama Fayad, Gregory Piatetky-Shapirio "The KDD Process for Extracting Useful Knowledge from Volumes of Data"ACM, vol 39,pp-27-34,1996.

[9] http:/www.nlm.nih.gov/mesh

[10] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," in Proc. International Conference on Management of Data, May 2000, pp. 1-12.

[11] M. J. Shaw, C. Subramaniam , G. W. Tan and M. E. Welge, “Knowledge management and data mining for marketing”, Decision Support Systems, v.31 n.1, pages 127-137, 2001.
[12] G. Saporta, “Data mining and official statistics”, In Proceedings of Quinta Conferenza Nationale di Statistica, ISTAT, Roma, 15 Nov. 2000.
[13] D. Malerba, F. Esposito and F.A. Lisi, “Mining spatial association rules in census data”, In Proceedings of Joint Conf. on "New Techniques and Technologies for Statistcs and Exchange of Technology and Know-how”, 2001.
[14] G. Serban, I. G. Czibula, and A. Campan, "A Programming Interface For Medical diagnosis Prediction", Studia Universitatis, "Babes-Bolyai", Informatica, LI(1), pages 21-30, 2006.
[15] Akash Rajak and Mahendra Kumar Gupta, “Association Rule Mining: Applications in Various Areas”, International Conference on Data Management.