Multi-Level Association Rule Mining: A Review

Sonia Setia [1], Dr. Jyoti[2]

[1]*Computer Engineering, Echelon Institute of Technology, Faridabad, India*

[2]*Computer Engineering, YMCAUST, Faridabad, India*

Abstract : *Association rule mining is the most popular technique in the area of data mining. The main task of this technique is to find the frequent patterns by using minimum support thresholds decided by the user. The Apriori algorithm is a classical algorithm among association rule mining techniques. This algorithm is inefficient because it scans the database many times. Second, if the database is large, it takes too much time to scan the database. For many cases, it is difficult to discover association rules among the objects at low levels of abstraction. Association rules among various item sets of databases can be found at various levels of abstraction. Apriori algorithm does not mine the data on multiple levels of abstraction. Many algorithms in literature discussed this problem. This paper presents the survey on multi-level association rules and mining algorithms.*

Keywords - *Data mining, Association rule mining algorithm, minimum support threshold, multiple scan, multi-level association rules.*

I. INTRODUCTION

Association rule mining identifies associations among database attributes and their values. It is a pattern-discovery technique which generally searches for associations among attributes present in the database. For an example: "if a customer buys a conditioner, then he also probably buys shampoo (in the same transaction)", the rule can be written as: {conditioner} → {shampoo}.

Now a days, Association rule mining is primarily used by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among various factors such as price, staff skills, competition, customer demographics and economic indicators. It also impacts on sales, customer satisfaction, and corporate profits.

In this paper, we surveyed the most recent existing association rule mining techniques. The organization of the rest of the paper is as follows. Section 2 provides the basic concepts and nomenclature for association rules to facilitate the discussion and describes the well known algorithms. Section 3 presents the survey on previous related work done on multi level association rule mining techniques. Finally, Section 4 concludes the paper.

II. BASIC CONCEPTS

Association rule (AR) is commonly understood [1] as an implication in the form of $X \rightarrow Y$ in a transaction database where, X and Y are the subset of item sets in the database and $X \cap Y \neq \phi$. There are two basic measures for association rules. They are:

1. Support(s) - Support(s) can be defined as the fraction of records that contain XUY to the total number of records in the transaction set D.

2. Confidence(c) – Confidence(c) can be defined as the fraction of records that contain XUY to the total number of records that contain X.

Generally, an association rule mining process completes in following steps:

- Support Threshold is the predefined by the user.
- The set of potentially large k-item sets is generated by 1-extensions of the large (k -1) - item sets generated in the previous iteration, where k is the length of itemset. These itemsets are called candidate itemsets.
- Support is calculated for these candidate k-item sets.
- Item sets having support less than the predefined support threshold are discarded and the remaining item sets are called large k-item sets.

This process is repeated until no larger item sets are found.

For many cases, it is difficult to discover association rules among the objects at low levels of abstraction. Association rules among various item sets of databases can be found at various levels of abstraction. The rules which are generated by mining the data at multiple levels of abstraction are called Multi-level association rules. Association rules discovered at high levels represent common sense knowledge. Multi-level association rules can be discovered efficiently using concept hierarchies

which satisfy minimum support-confidence framework. In general, top-down progressive deepening method is used for mining Multi-level association rules, where support is counted to generate frequent item sets at each level starting from the level 1 and moving downward in the concept hierarchy, until no more concept items can be found. Thus one might be interested in discovering frequent item sets composed of items which themselves form taxonomy. Many algorithms in literature discussed this problem.

Multilevel databases use encoded transaction table driven by using concept hierarchy information instead of the original transaction table. Let's take an example of sales transaction database as shown in Table 2.1 which illustrates the items purchased for each transaction.

Table 2.1: Sales Transaction Database

Trans_Id	**Items purchased**
T100	Amul double toned milk, White bread
T200	Coca cola Drink
T300	Real mix juice
T400	Tata black tea
…	…

The concept hierarchy formed for these items is shown in Fig 2.1:

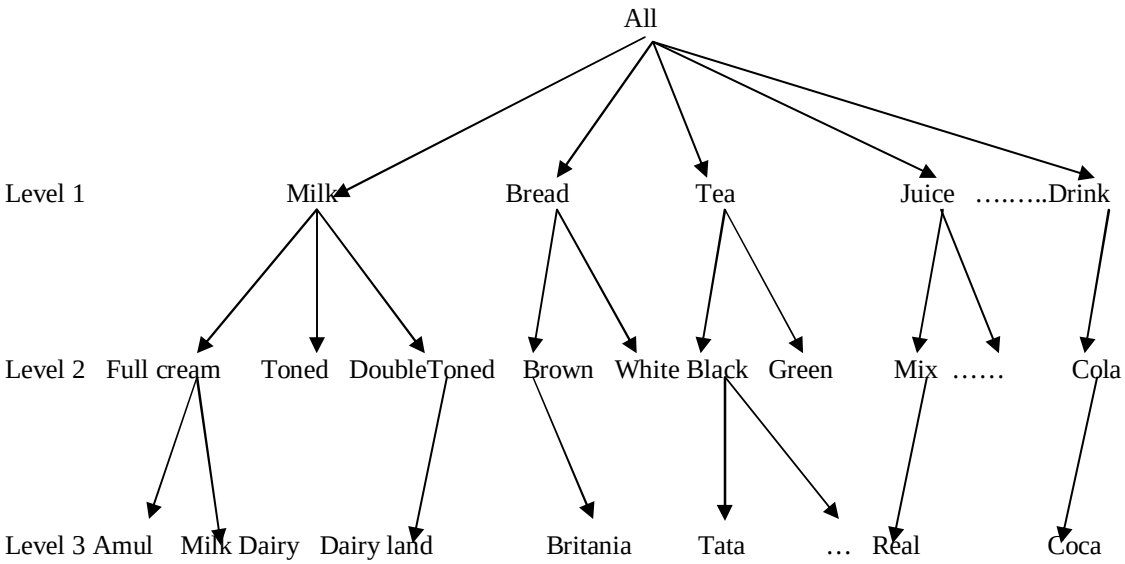

Fig 2.1: Concept hierarchy

Thus, the encoded transaction table can be derived by using this concept hierarchy. Encoded string represents the position of item in the concept hierarchy and it requires less number of bits. By using encoding scheme, more items can be merged due to their identical encoding, which further reduces the size of the encoded transaction table. Thus, it is beneficial to use encoded transaction table instead of original transaction table.

For the above said example, derived encoded transaction table from the concept hierarchy shown in fig 2.1 is given in Table 2.2.

Table 2.2: Encoded transaction table

Level	Code	Items
Level 1	1**	Milk
	2**	Bread
	3**	Tea
	4**	Juice
	…	…
Level 2	11*	Full cream
	12*	Toned
	13*	Double toned
	21*	Brown
	22*	White
	…	…
Level 3	111	Amul
	121	Mother dairy
…	…	…

III. RELATED WORK

In [3], R.Srikant et al. introduced the problem of mining generalized association rules. Given a large database of transactions, where each transaction consists of a set of items, and a hierarchy of the items, they find associations between items at any level of the hierarchy.The author quoted that earlier work on association rules did not consider the hierarchy of the transaction, and restricted the items in the association rules to only leaf-level items in the hierarchy. Author also added that an obvious solution to the problem is to consider all the items in the original transaction as well as all the ancestors of each item in the original transaction. Then any of the earlier algorithms can be used for mining association rules on these extended transactions to mine generalized association rules. The author described two algorithms, Cumulate and EstMerge, which run 2 to 5 times faster than Basic.

In [2]Jiawei Han et al. developed a top down progressive deepening method for mining multiple level association rules from large transaction databases by extending some existing association rule mining techniques. A group of variant algorithms are proposed based on the ways of sharing intermediate results, with the relative performance tested on different kinds of data. Relaxation of the rule conditions for finding "level-crossing" association rules is also discussed in this paper.

Earlier work on multi level association rule mining ignore the fact that the taxonomies of items cannot be kept static while new transactions are continuously adding into the original database. To reflect the database change with taxonomy evolution, transaction update is a crucial task.

In [11],Tseng et al. examined this problem and proposed a novel algorithm, called IDTE, which can incrementally update the discovered generalized association rules with the taxonomy evolution with new transactions insertion to the database. Empirical evaluations show that this algorithm can perform well even in large amounts of incremental transactions and high degree of taxonomy evolution.

In [4], R. S. Thakur et al. proposed a new algorithm for mining multi-level association rules in large databases. It uses concept of counting inference approach that allows performing as few support counts as possible. This new method reduces database scan at each concept level than the other existing algorithm for multi-level association rule mining from large databases.

In [5], Yin-bo Wan et al proposed a novel method to improve the situation by analyzing the rules mined from primitive concept level to obtain multilevel rules. The proposed method also supports dynamic concept hierarchies.

In [6], Pratima Gautam et al. proposed a Boolean Matrix based approach to discover frequent item sets. It adopts Boolean relational calculus to discover maximum frequent item sets at lower level. It scans the database at every level to generate the association rules. In this paper, Apriori property is used to prune

the item sets. It used top-down progressive deepening method and also used Boolean logical operation to generate the multilevel association rules.

In [7], authors employed partition and Boolean concepts, multiple level taxonomy and different minimum supports to find association rules in a given transaction data set. The model works well with problems involving uncertainty in data relationships, which are represented by Boolean concepts. The proposed algorithm can thus generate frequent item sets level by level and then derive association rules from transaction dataset. It also examines method of partitioning to limit the total memory requirement.

In [8], Virendra Kumar Shrivastava et al. proposed a method for discovery of multi-level association rules from primitive level FP-tree in order to reduce main memory usage and make the execution faster. This proposed method constructs FP-tree from FP-tree of primitive level. To generate frequent pattern in multi-level it uses COFI-tree method which reduces the memory usage in comparison to ML_T2L1. Therefore, it can mine larger database with smaller main memory available and it also runs fast. This method uses the non recursive mining process and a simple traversal of the COFI-tree, a full set of frequent items can be generated. It also uses an efficient pruning method that is used to remove all local non frequent patterns, leaving the COFI tree with only local frequent items. It reaps the advantages of both the FP growth and COFI.

Shin-Mu Vincent Tseng [9] proposed an intelligent and efficient technique for mining association rules with constrained categories in large databases. The proposed technique consists of two components: a data mining algorithm and a database analyzer called Pre-miner. The proposed mining algorithm can efficiently discover the association rules between the data items in a large database based on the dynamic constraints specified by the users. In particular, only one scan of the whole database is needed for each query. Hence, the high repeated disk overhead can be reduced significantly.

In [12], authors proposed an efficient method for mining association rules at multiple levels. This method is based upon the Boolean matrix approach and extends the algorithm proposed by the authors of [6] which is also based on similar approach. Authors quoted that the drawback of [6] is that it has to scan the database at every level for generating large item sets. Whereas, the proposed algorithm scans the database once and generates large item sets for all levels in one go. While scanning the database, it generates large 1-item set for all levels in parallel. There is no need to scan the database further to produce the itemsets for lower levels. It uses Boolean logical AND operation to generate large k-item sets for each level.

In [13], a new method is proposed to discover multilevel association rules which is based on spatial topology relationship between different objects. The author used DE-91M theory to find the spatial relationship expressions. There are some problems existing in multi-level association rules mining:

- When the minimum support set is very high, the rules are less and there is lack of representative.
- When the minimum support set is too low, meaningless association rules may be produced.

In order to enhance efficiency of mining, the concept of 'meta-rule' is introduced in this paper. Thus efficiency of the method is enhanced by the MBR Test and the Meta rule. Experimental result shows that the method is valid and more efficient.

Most of the approaches mentioned above focus on binary valued transaction data but such in real-world applications usually consists of quantitative values. Fuzzy data mining algorithms have thus been proposed for handling quantitative transactions and mining fuzzy association rules. Moreover, earlier work on Association rule mining is based on the assumption that users can specify support threshold to mine their databases. However, setting the support threshold is a difficult task for users. This can hamper the widespread applications of ARM algorithms. In view of the mentioned insufficiencies, authors [14] proposed a fuzzy multi level mining algorithm to automatically compute the support threshold for each item. In first phase, it computes minimum support threshold for each item in database according to user's requirement and in second phase, starting with multiple minimum support of items , it extracts knowledge implicit in quantitative transactions. This work results in two benefits :

- Computing the minimum support for each item in database
- Making a system automation

In [15], authors proposed a Multi-Objective Multi Level Genetic-Fuzzy Mining (MOMLGFM) algorithm for mining a set of non-dominated membership functions for discovering multilevel fuzzy association rules. Firstly, it encodes the

membership functions of each item class into a chromosome according to the given taxonomy. Then it calculates the two objective functions for each chromosome. The first function is calculated by summing the large l-itemsets of each item in different concept levels, and the second one is the suitability of membership functions. By using these two objective functions, fitness value of each individual is then evaluated. After the termination of evolution process, various sets of membership functions could be used for deriving multiple-level fuzzy association rules according to the different criteria of the decision maker. Experimental result shows the effectiveness of the algorithm.

A fusion model [16] is introduced to discover multi-level fuzzy association rules. In this work, cumulative probability distribution approach (CPDA) is integrated with multi-level taxonomy concepts to mine fuzzy association rules. The proposed model generates large itemsets level by level and mine multi-level fuzzy association rule which leads to find more informative and important knowledge from transaction dataset.

Clustering classification based new method [17] is proposed to mine multi-level association rules. The proposed method is the integrated approach of hierarchical concept, generalization sets processing, and SOFM neural network generalization. This method also generates an internal threshold so no need to specify the minimum support threshold. This method not only reduces candidate itemsets, but also reduces the database scan counts, to make the ARM more valuable and practical significant.

IV. CONCLUSION

Association rules are very useful in applications going beyond the standard market basket alanysis. In this paper, we surveyed the list of existing multi-level association rule mining techniques. We have shown here various multi-level association rule mining algorithms used to find frequent items in a given transaction of database.

REFERENCES

[1] R. Agrawal., T. Imielinski., and A. Swami., "Mining association rules between sets of items in large databases". In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216, Washington, DC, May 26-28 1993.

[2] J. Han., and Y. Fu., "Mining Multi level Association Rules in Large Databases". IEEE Knowledge and Data Engineering, Vol 11, pp 798-805, 1999.

[3] R. Srikant., and R. Agrawal., "Mining Generalized Association Rules". Proc. Of the 21th International Conference on Very Large Data Bases, pp. 407-419, 1995.

[4] R. S. Thakur., R. C. Jain., and K. R. Pardasani., "Fast Algorithm for mining multi-level association rules in large databases". Asian Journal of International Management Vol. 1, pp. 19-26, 2007.

[5] Y. Wan., Y. Liang., and L. Ding., "Mining multilevel association rules with dynamic concept hierarchy". In proceedings of the seventh international conference on machine learning and cybernetics, kunming, pp. 12-15, July 2008.

[6] Pratima Gautam., and K. R. Pardasani., "A Fast Algorithm for Mining Multilevel Association Rule Based on Boolean Matrix". (IJCSE) International Journal on Computer Science and Engineering, Vol. 2, pp. 746-752, 2010.

[7] Pratima Gautam., and K. R. Pardasani., "Efficient Method for Multiple-Level Association Rules in Large Databases". Journal of Emerging Trends in Computing and Information Sciences, Vol. 2, December 2011.

[8] Virendra Kumar Shrivastava., Parveen Kumar., and K. R. Pardasani., "Discovery of Multi-level Association Rules from Primitive Level Frequent Patterns Tree". International Journal of Computing Science and Communication Technologies, Vol. 3, July 2010.

[9] Shin-Mu Vincent Tseng." An Efficient Method for Mining Association Rules

with Item Constraints". Technical Report No. CSD-99-1089 Berkeley, CA94720.

[10] Y. Li., and R.Gopalan., "Effective Sampling for Mining Association Rules", Lecture Notes in Computer Science, Vol. 3339, pp. 391 – 401, Jan 2004.

[11] M. Tseng., W. Lin., and R. Jeng., "Maintenance of Generalized Association Rules UnderTransaction Update and Taxonomy Evolution". Lecture Notes in Computer Science, Vol. 3589, pp. 336–345, Sep 2005.

[12] Sonia Setia, Dr. Jyoti, "Efficient Mining of Multiple-level Association Rules using Boolean Matrix" Proceedings of the 7th International Conference on Advanced computing and communication technologies (ICACCT- 2013) ISBN: 978-93-83083-38-1, 2013.

[13] Tao Feng, Zhiyong Zeng, XuChao Wu, Rui Liu, Lianxiong Gao "Discovery of multi-level spatial association rules based on DE-9IM" 978-1-4244-5326-9/10 ©2010 IEEE.

[14] Mahmoudi E.V., Vahid Aghighi, M.N. Torshiz, M. Yaghoobi, "Mining Generalized Fuzzy Association Rules via Determining Minimum Supports" 19th Iranian Conference on Electrical Engineering, 2011.

[15] Chun-Hao Chen, Chi-Hsuan Ho, Tzung-Pei Hong, Wei-Tee Lin, "MOGA for Multi-Level Fuzzy Data Mining", IEEE International Conference on Granular Computing, 2012.

[16] Jr-Shian Chen, Fuh-Gwo Chen, Jen-Ya Wang, "Enhance the Multi-level Fuzzy Association Rules based on Cumulative Probability Distribution Approach",13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 978-0-7695-4923-1/12 © 2012 IEEE.

[17] Huang QingLan, Duan LongZhen, "Multi-level association rule mining based on clustering partition" , Third International Conference on Intelligent System Design and Engineering Applications, 978-0-7695-4923-1/12, © 2012 IEEE