Effective and Innovative Approaches for Comparing Different Multilevel Association Rule Mining for Feature Extraction: A Review

Alisha S. Patel
Research Scholar
Department of Computer Science
Parul Institute of Technology, Vadodra, Gujarat

Mohit Patel
Assistant Professor,
Department of Computer Science
Parul Institute of Technology, Vadodra, Gujarat

ABSTRACT
Today multilevel association rule mining is an emerging field in data mining. Its main goal is to find hidden information in or between levels of abstraction. It is mainly used for decision making for large data. It focuses on the customer relationship management. Apriori algorithm is mainly used for the multilevel association rule mining. Producing large number of candidate item sets and multiple scanning databases is main shortage of the Apriori algorithm. Because of the multilevel association execution time is reduced and throughput increase in new methods.MRA algorithm using Bayesian probability, concept hierarchy ,COFI-tree method, dynamic concept hierarchy are used for increase performance of multilevel association rule mining.

Keywords
Data mining, Association Rule Mining, Multilevel Association Rule Mining, Concept Hierarchy.

1. INTRODUCTION
Data mining is the process of discovery of the knowledge from the large amount of the data or databases. Data mining functions contain clustering, classification, prediction, association and multilevel association rules [8] .One main application area of data mining is to discover potentially useful information from transaction databases and large amount of data. Data mining is an interdisciplinary subfield of computer science, it is a process of discovering patterns in large data sets or database involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems[2]. The main goal of the data mining process is to extract information from a data set and transform it into an understandable structure for future use.

The data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records or cluster analysis, anomaly detection and dependencies that calls association rule mining. These patterns can then be seen as a kind of summary of the input data, and used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system.

Today lots of data is being collected and stored in the warehouse, which is called data warehouse. Raw ideas from machine learning, pattern recognition, statistics, and database systems data mining is originate. Data mining is also referred as the core processes of Knowledge Discovery in Database (KDD) [15].

2. ASSOCIATION RULE MINING
Association rule is the process of finding the relationship among the items of database. It first introduced by Agrawal in 1993 to identify relationship between the items of database [11].It is widely used technique in data mining. It is used for decision making purpose and improvement of business [9]. The purchasing of one product when another product is purchased represented by an association rule. It is the relationship between two item sets *X* and *Y*. It can be denoted by X $\rightarrow$ *Y*. Means whenever X occurs then Y also occurs. For example, the rule {Shampoo $\rightarrow$Conditioner} [10].

Example 1: A shopping mall has diwali specials for which advertising supplements are created for the local newspaper. When an item, such as dove shampoo, has been designated to go on sale, management determines what other items are frequently purchased with dove shampoo. They find that dove conditioner is purchased with dove shampoo 40% of the time and paracute oil is purchased with it 30% of the time. Based on these associations, special displays of oil and conditioner are placed near the dove shampoo which is on sale. They did not put sale on these items. By doing this overall sales volume will be increase and by taking advantage of the frequency with which these items are purchased together [8].A basic statement of the association rule problem proposed by Agrawal is:

Definition 1: Let I = {I1, I2, … , In} be a set of n distinct attributes which are called *literals*.

Let D be a database, where each tuple T has a unique identifier (TID), and contains a set of items such that T⊆I. An association rule is an implication of the form X⇒Y, where X, Y⊂I, are sets of items called itemsets, and X! Y=φ. Here, X is called antecedent, and Y consequent [8]. The basic two measures of association rule mining are Support and Confidence.

Support (A) = Proportion of transaction in the data set that contain the itemset to the total number of transactions.

Confidence (A$\rightarrow$B) =Ratio of total number of transaction with all the items to the number of transaction with the A item set.

An association rule mining process consist the following steps:

- User will predefine the threshold.

- Large k-item sets are generated by 1-extensions of the large (k -1) -item sets which is generated in the previous iteration, where k defines length of itemset, which are called candidate itemsets.

- For these candidates k-item sets support is calculated.

3. MULTILEVEL ASSOCIATION RULE MINING

Multi-level association rules contain mining of the same level and across levels. Its complexity is higher than the single level association rule mining. Multi-level association rules is in the form of: $X \Rightarrow Y$. Among them, $X \subset I$, $Y \subset I$, $X \cap Y = \Phi$, and arbitrary items of Y are not items' ancestors in X. The support degree is: support$(X \Rightarrow Y)$ = support$(X \Rightarrow Y)$; The confidence is: confidence $(X \Rightarrow Y)$ = support $(X \cup Y)$/ support$(X) \times$ 100%[1].

Different rules that are generated by mining the data at multiple levels of abstraction are called Multi-level association rules. In multilevel association rule, organizations want to evaluate association between itemset at a different level of hierarchies because items often form hierarchies as shown below. For example Organization needs to find how many customers buy jeans with shoes and t-shirt with sandals.
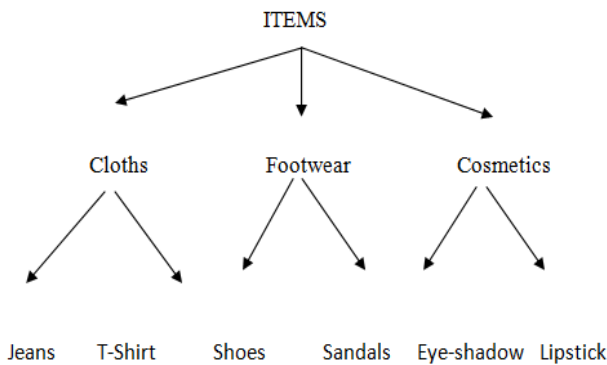

Figure 1: Multilevel hierarchy

Rules can help organizations to make promotional strategies and help to increase the sales and setting the plans for future. Multi-level association rule mining is a top-down approach. In this approach at the highest level some frequent item set is found as we do in single level association rule mining , then move towards the next level (lower level) to find out the frequent item set at that level of hierarchy and continue moving to the lowest level until no more frequent item set can be found. For example

- Find out the frequent item set and then mine strong association rule:

{Cloths, Footwear} {Cosmetics} {30%, 60%}

- After that find out the lower level rules:

{Jeans, T-Shirt} {Cosmetics} {5%, 50%}

There are two types of support on which multiple level association rule mining can work which are- Uniform support and Reduced support.

1. Uniform Support: Constant minimum support threshold is used at every level of hierarchy in this. Here there is no need to evaluate itemsets containing items whose ancestors do not have minimum support. There should be appropriate minimum support threshold. When minimum support threshold is too high then lower level associations will lose and if too low then it will end up in generating too many uninteresting high level association rules.

2. Reduced Support: At the lower levels reduced minimum support is used. Search strategies are as above:

I. Independently level by level

II. Filtering across the levels by k-itemset

III. Filtering across the levels by single item

IV. Controlled level filtering by single item.

4. LITERATURE SURVEY

In [1] this paper author proposed hierarchical concept. It uses the data of the generalization sets processing. It uses SOFM neural network generalization into the database after the transaction, using way of introducing an internal threshold. There is no need to set the minimum support threshold, for that internal threshold is introduced. This method mine efficiently for single-layer and cross-layer association rules, but also the association rules is new ,easy to understand and meaningful. ML_T2L1 algorithm and the Cumulate algorithm are used for the multilevel association rule. In this paper author proposed clustering based on the division of multilevel association rules mining, through clustering division affairs database, to make frequent items relatively concentrated data in each subset, and introducing the internal threshold way to need not set the minimum support threshold. Using this database scanning time will be decrease. SOFM neural network is used for the clustering partition. Compared with other multi-level association rules algorithm, proposed algorithm can effectively reduce the number of scanning database, greatly reduce the candidate itemsets generation and improve the efficiency of association rules mining of the same level and across levels. From the result we can say that this new algorithm works best than the apriori algorithm. New algorithm takes less time in execution. New algorithm performs more effective, efficiently and feasible. In [2] author proposed fast algorithm for solving the problem of multilevel association rule mining. They present an efficient version of apriori algorithm for finding maximum frequent itemset at lower level of the abstraction. To reduce the execution time and increase throughput one new approach is proposed by them which is SC-BF multilevel algorithm. For counting the frequent itemset at each level support count table and bit from table is used. Some advantages of this algorithm are smaller set of high quality rules directly from the dataset, grouping of the items in each level, to maintain the large set of database, it reduce the number of the records in dataset. At the top most level it first find out frequent item sets and then deepens the process into their frequent descendants at lower concept level. This method use progressive deepening method and concept of the reduce support. At branch level data were grouped so it takes less memory to store and reduce the execution time. It performs efficiently and more accurately. SC-BF multilevel algorithm is based on the occurrences of the item and then by performing the logical AND operation. In this proposed algorithm level by level generation of the large dataset is occurs and time will be saved without degrading the accuracy of the general approach. This will eliminate repeated scanning of transaction database. In [3] this paper author proposed frequent and infrequent association rule. For in-depth association mining multilevel association rule is used. Author introduced a new model which allows the user to specify multiple minimum supports to reflect different natures and

frequencies of items. It helps user to generate rare item sets. For the development of multilevel propositional logic based knowledge discovery involve steps like data pre-processing, Propositional Logic for Coherent Rule Generation, Multi-Level concept hierarchy and Performance measure on generated rules. The proposed mining algorithm can generate large item sets level by level and then derive concept multilevel association rules from transaction dataset. Proposed algorithm can derive the multiple-level association rules under different propositional logic in a simple and effective way. Proposed multilevel association rule reduce the risks associated with using an incomplete set of association rules for decision making. The multilevel coherent rules mining framework is used for its ability to discover rules that are both implicational and complete according to propositional logic from a given data set at various conceptual levels. In [4] this paper author proposed dynamic concept hierarchies to improve the situation by analyzing the rules mined from primitive concept level to obtain multilevel rules. Proposed method supports multiple hierarchies which can be dynamically constructed according to the input of user knowledge. Concept of FP'-tree and its corresponding constructing method is introduced. They use FP(l)-tree to denote the FP'-tree at the concept level l. Multilevel association rules at specific levels algorithm is designed to generate the candidate association rules at user specified level. Mining efficiency will increase by merging and grouping the rules at atomic level. This approach can support dynamic hierarchies based on different views of organizing items, which allows different users get their desired association rules from customized point of view. In [5] this method author presents Multilevel Relationship Algorithm. Apriori algorithm is used for generating frequent item set so it is call Multilevel Relationship algorithm. Three stage of working is there. In first two stages it perform MRA and after that in the third stage it perform Bayesian probability to find out the dependency & relationship among different shops, pattern of sales & generates the rule for learning. Efficiency of original Apriori algorithm has been increased due to multilevel architecture. From the result evaluation is that MRA performs better than the Apriori algorithm. MRA is always takes less time than the Apriori algorithm. As number of support increases it decreases the time requirement for both MRA and Apriori algorithms. MRA performs better than the Apriori algorithm towards improvement of mining association rule. In [6] author proposed methods to reduce main memory usage and make execution faster author proposed a method for discovery of multi-level association rules from primitive level FP-tree. The proposed method used hierarchy information encoded transaction table instead of the original transaction table. Proposed method constructs FP-tree from FP-tree of primitive level. COFI-tree method reduces the memory usage in comparison to ML_T2L1 for generating frequent pattern in multilevel. This method runs fast and can mine larger database with smaller main memory available. Proposed method reduced main memory usage and execution complexity. In [7], author proposed RFID data represented as a concept hierarchy. Author proposed two aims in this paper. First is time generalization to eliminate data duplication. Second is to reduce the complexity of rule generation. Concept hierarchies introduce semantic structure which can be exploited to reduce the complexity of data, focusing mining on potentially interesting mining queries. From the experimental result conclusion is that the more the inclusion information between nodes is increasing, the more the number of rule generating by generalization is decreasing considerably. Therefore duplicated rule generation will be

eliminated and decrease the cost of rule generation by concept hierarchy at the time of the association rule generation between nodes. Complexity of the rule generation is reduced. Due to filtering of the data time is reduced and query speed increase.

4.1 Comparative analysis
Table 1: Comparative Analysis

Title	Method	Analysis
"Multi level association rule mining based on clustering partition[1]"	Clustering Partition(SOFM)	In this paper, SOFM neural network and internal threshold used. Using SOFM we can reduce the candidate item sets and time also.
"To improve Association Rule Mining using New Technique: Multilevel Relationship Algorithm towards Cooperative Learning[5]"	MRA(Bayesian probability)	In this paper, Multilevel Relationship Algorithm used. Comparison of Apriori and MRA is used for different shops, where MRA is more efficient and takes less time.
"Algorithm for Efficient Multilevel Association Rule Mining[2]"	SC-BF multilevel algorithm	In this paper, SC-BF multilevel algorithm is used, based on the occurrences of the item and then by performing the logical AND operation it will reduce time and scanning time then apriori.
"Mining Multilevel Association Rules on RFID data[7]"	Hash Method, Concept Hierarchy	Complexity of rule generation decrease, Reduce time and query speed
"Discovery of Multi-level Association Rules from Primitive Level Frequent Patterns Tree[6]"	FP-Tree	COFI-tree method reduces the memory usage in comparison to ML_T2L1,Run fast, Reduce memory usage and execution complexity
"A Novel Method of Mining Association Rule with Multilevel Concept Hierarchy[3]"	Multilevel concept hierarchies	The multilevel coherent rules mining framework is used for its ability to discover rules that are both implicational and

		complete according to propositional logic from a given data set at various conceptual levels
"Mining multilevel association rules with dynamic concept hierarchy[4]"	FP Growth(FP - Tree),dynamic concept hierarchy	Improves mining efficiency at atomic level, allow users to get their desire association rule using concept hierarchy

5. CONCLUSION

The traditional multi level association rule mining based on the Apriori algorithm is not applicable to a large database. Association rules are very useful in applications going beyond the standard market basket analysis. In this paper, we surveyed and describe the list of existing multi-level association rule mining techniques. FP-tree is performing better then the apriori algorithm.

Using concept hierarchies for the RFID data storage space will be minimize and increase search capacity. Using COFI-tree method which reduces the memory usage in comparison to ML_T2L1.It is feasible and effective to use multilevel association rule mining using clustering division. MRA (Multilevel Relationship Algorithm) performs better than the Apriori algorithm towards improvement of mining association rule. Using SOFM algorithm scanning of the database will be reduced and time will also reduce as compare to the apriori algorithm. In future different unsupervised learning algorithm can be used for multilevel association rule mining with combination of alternatives of apriori algorithm.

6. REFERENCES

[1] Huang QingLan, Duan LongZhen, "Multi-level association rule mining based on clustering partition", IEEE(2013).

[2] Pratima Gautam, Dr. K. R. Pardasani , "Algorithm for Efficient Multilevel Association Rule Mining", IJCSE(2010).

[3] S.Prakash, M.Vijayakumar, " A Novel Method of Mining Association Rule with Multilevel Concept Hierarchy", IJCA(2011).

[4] YIN-BO WAN, YONG LIANG, LI-YA DING, "Mining Multilevel Association Rules With Dynamic Concept Hierarchy", IEEE(2008).

[5] Deepak Vidhate, Dr. Parag Kulkarni, "To improve Association Rule Mining using New Technique: Multilevel Relationship Algorithm towards Cooperative Learning" ,IEEE(2014).

[6] Virendra Kumar Shrivastava1, Parveen Kumar2, K. R. Pardasani3, "Discovery of Multi-level Association Rules from Primitive Level Frequent Patterns Tree", IJCSCT(2010).

[7] Younghee Kim, Ungmo Kim, "Mining Multilevel Association Rules on RFID data", IEEE (2009).

[8] Margaret H. Dunham, Yongqiao Xiao Le Gruenwald, Zahid Hossain, "A survey of association rules".

[9] Nikunj H. Domadiya, Udai Pratap Rao , "Hiding Sensitive Association Rules to Maintain Privacy and Data Quality in Database", IEEE(2012).

[10] Sonia Setia 1, Dr. Jyoti2, "Multi-Level Association Rule Mining: A Review", IJCTT(2013).