

# Introduction to Multilevel Association Rule and Its Methods

Jinal J. Shah<sup>1</sup> Lokesh P. Gagnani<sup>2</sup>

<sup>1</sup>ME Scholar <sup>2</sup>Assistant Professor  
<sup>1,2</sup>IT, Kitrc-Kalol, Gandhinagar-382721

**Abstract**— Association rule mining is a popular and well researched method for discovering interesting relations between variables in large databases. In this paper we introduce the concept of Data mining, Association rule and Multilevel association rule with different algorithm, its advantage and concept of Fuzzy logic and Genetic Algorithm. Multilevel association rules can be mined efficiently using concept hierarchies under a support-confidence framework.

**Key words:** Multilevel Association rule, Support, Confidence, Genetic-Fuzzy algorithm, Fuzzy set

## I. INTRODUCTION

### A. Data mining:

Data mining is an interdisciplinary subfield of computer science is the computational process of discovering patterns in large data sets ("big data") involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

### B. Technologies in Data mining:

Rapid advance in data capture, transmission and storage, Data Mining is Associated with many other technologies that will help to implement new and innovative ways to mine the after- Market value of their vast stores of detailed & massy data. The most commonly used techniques in data mining are:

#### 1) Artificial Neural Networks:

Non-linear predictive models that learn through training and resemble biological neural networks in structure.

#### 2) Classification:

Classification is a data mining technique used to predict group membership for data instances. Genetic algorithms: Optimization techniques that proceed as genetic combination, mutation, and natural selection. Mostly used in classification & association rule extraction.

#### 3) Association Rules:

It rely on finding dependencies between features which are situated in set of data. This technique tries to find rules which fulfill conditions that user set. They can be used in finding sets of the most frequently together visited web-sites.

#### 4) Clustering:

It concerns finding in set of elements certain subsets. Every subset has elements which are similar between themselves but they are different from the rest of the data.

## II. ASSOCIATION RULE MINING

Association Rule mining is the scientific technique to dig out interesting and frequent patterns from the transactional, spatial, temporal or other databases and to set associations , relations or correlations among those patterns (also known as item sets) in order to discover knowledge or to frame

information[1]. Association rules can be applied in various fields like network management, Basket data analysis, catalog design, clustering, classification, marketing etc. Association rules establishes the relationship between different variables to analyses the present situation. For e.g. to find the relationship between the various items sold at a shopping mall the association rule can be applied on the huge amount of data recorded by the Shopping mall.

### A. Support:

Support is a basic measure related to probability and set theory. It is defined as the fraction of transactions in the database which contain all items in a specific rule [1]. This can be written as:  $Supp(X \Rightarrow Y) = Supp(XUY) = |xy| / |D|$ .

Where  $|xy|$  is the number transactions (item set) which contain both X and Y –i.e., the probability of (x, y)–and  $|D|$  represents the total number of transactions (item set) in the database. Minimum support threshold is usually specified in generating the association rules which select only the most frequent items in the database [11].

### B. Confidence:

Another measure of the association rules is confidence. This is the strength of the implication of a rule and can be represented as a ratio between the transaction numbers, including X and Y and those including X, which can be written as:

$$Conf(X \Rightarrow Y) = Supp(XUY) / Supp(X) = |xy| / |x|.$$

Where  $|x|$  is the number of transactions (itemset) containing X.[11] For e.g. the rule {Computer , Printer} - {UPS} found in the sales data of a mall would indicate that if a customer buys Computer and Printer together, he or she would definitely also buy UPS.

The rule is

$$\{\text{computer, printer}\} \rightarrow \{\text{UPS}\}[1]$$

Transaction Id	Item set
1	{computer, printer}
2	{computer,printer,Ups}
3	{computer}
4	{computer,printer,Ups}
5	{Ups}
6	{computer,printer,Ups}

Table 1:

This implies that if customer buy Computer and Printer. He tends to buy UPS also. Out of 6 transaction 3 transactions support this rule .In 3 records all the three items are brought together. So the Support of rule denoted as  $Supp(A)$  is proportion of transaction in the data set that contain the itemset to the total number of transactions. In the above example, the itemset {Computer, Printer , UPS} has a support of  $3/ 6 = 0.5$  since it occurs in 50% of all transactions (3 out of 6 transactions).

The Confidence of a rule (denoted as  $conf(A B)$ )=Ratio of total number of transaction with all the items to the number of transaction with the A item set . for e.g Computer and Printer are purchased 4 times and out of 4

transactions UPS is purchased three times with Computer and Printer i.e. A so the  $\text{conf}(A,B) = \frac{3}{4} = .75$  i.e. 75%. So the association rule is the technique to set the relation between item sets to draw important conclusions. It sets the minimum support and confidence threshold and evaluates the frequent itemset and then uses the evaluated itemset to frame desired results. The general framework of ARM is shown in Fig 1[8].

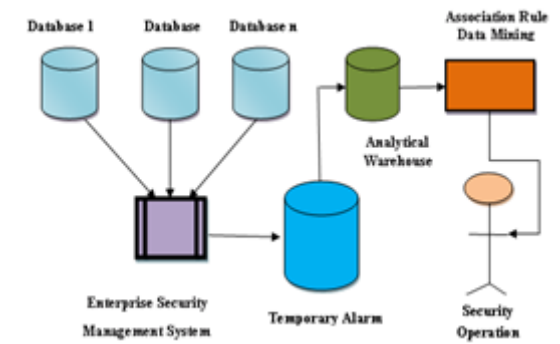


Fig. 1:

1) It classifies into following Categories:

- Boolean or quantitative associations
- Single dimension or multidimensional associations.
- Single level or multilevel associations

### III. MULTILEVEL ASSOCIATION RULE

Association rules generated from mining data at multiple levels of abstraction are called multiple-level or multilevel association rules. Multilevel association rules can be mined efficiently using concept hierarchies under a support-confidence framework. It is difficult to find strong associations among data items at low or primitive levels of abstraction due to the sparsity of data at those levels. Strong associations discovered at high levels of abstraction may represent commonsense knowledge. Data mining systems should provide capabilities for mining association rules at multiple levels of abstraction, with sufficient flexibility for easy traversal among different abstraction spaces. Rules regarding item sets at suitable levels could be relatively functional. It can help organizations to make promotional strategies and help enhancing the sales and setting the future plans. A basic approach to multi level association rule mining is top-down progressive deepening approach. In this approach first find out the frequent item set at the highest level as we do in single level association rule mining, then move to 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.

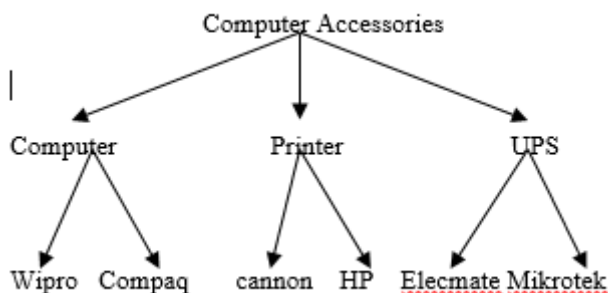


Fig. 2:

Rules regarding item sets at suitable levels could be relatively functional. A basic approach to multi level association rule mining is top-down progressive deepening approach. In this approach first find out the frequent item set at the highest level as we do in single level association rule mining, then move to 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 e.g. First find out the frequent item set and mine strong association rule : {Computer, Printer} → {UPS} {20%, 60%} Then find out the lower level rules : {wipro computer, cannon printer} → {UPS} {6%,50%}

Multiple level association rule mining can work with two types of support- Uniform and Reduced.

#### A. Uniform Support:

In this approach same minimum support threshold is used at every level of hierarchy. There is no need to evaluate itemsets containing items whose ancestors do not have minimum support. The minimum support threshold has to be appropriate. If minimum support threshold is too high then we can lose lower level associations and if too low then we can end up in generating too many uninteresting high level association rules. For e.g.

- At Level 1 Computer, Printer
- Minimum supp 5% [support – 10%]
- At Level 2 Wipro Computer, printer
- Minimum support 3% [support 7%]
- Wipro Computer, HP Printer [3%]

#### B. Reduced Support:

In this approach reduced minimum support is used at lower levels.

### IV. EXISTING APPROCHES FOR MLAR

There are many existing approaches to find multi level association rule. Some common methods are ML\_T1LA, ML\_TML1, ML\_T2LA, ML\_T2L1, Pincer search and Genetic fuzzy association rule. In the group of ML\_T\*L\* algorithm different tables and levels are used. For ex. In ML\_T1LA[10] only one transaction table is used and number of levels and main advantage of that method is that it avoids generation of new encoded table. In ML\_TML1[10], there are multiple encoded transaction tables are generated. The main advantage of it saves substantial amount of processing. In ML\_T2LA[10] there are two transaction tables. The main advantage of it - It avoids generation of group of new filtered transaction table. In ML\_T2L1[10] algorithm, It gathered a large itemsets on appropriate level the association rules can be mined. The main advantage of it is- seeks large itemset using digits and consumes less time. In pincer search[2] method, It can easily discover maximum frequent set. The main advantage of it is- It reduces I/O operation and processing faster. In Genetic-Fuzzy[3] method, it Searches membership functions suitable for multiple-level mining problems and then use the final best set of it. The main advantage of it is- Database shows effectiveness and stableness

## V. FUZZY LOGIC

Fuzzy logic is an enhance version of multivalve logic. However its uses are relatively different from their goals. Thus, the fact that fuzzy logic deals with approximate rather than precise modes of reasoning implies that, in general, the chains of reasoning in fuzzy logic are short in length and region does not play as important a role as it does in classical logical systems

### A. Fuzzy set:

Fuzzy sets are generalized sets which allow for a graded membership of their elements. Usually the real unit interval  $[0; 1]$  is chosen as the membership degree structure. Let  $X$  be a space of points, with a generic element of  $X$  denoted by  $x$ . Thus  $X = \{x\}$ . A fuzzy set  $A$  in  $X$  is characterized by a membership function  $f_A(x)$  which associates with each point in  $X$  a real number in the interval  $[0,1]$ , with the values of  $f_A(x)$  at  $x$  representing the "grade of membership" of  $x$  in  $A$ . Thus, the nearer the value of  $f_A(x)$  to unity, the higher the grade of membership of  $x$  in  $A$ .

## VI. NEED OF FUZZY ASSOCIATION RULE MINING

An intelligent system solves domain specific problems. Association rule mining is basically of two types. One is traditional or crisp association rule mining and the other one is fuzzy association rule mining. Classical association rule mining make use of Boolean logic to translate numerical attributes into binary attributes by facilitate of sharp crunchy partitions .In which valuable data may become conflicting over these sharp partitions [12].

Another difficulty with classical association rule mining is , here the user has the present minimum support value for the minimum purpose and as we know we humans are mistake prone. Any wrong settings of minimum support could end up in the incorrect result. So, it is a very difficult task of surrounding with the accurate minimum support rate by hand. That is why classification association rule is more time consuming and fewer accurate process [12].

## VII. GENETIC ALGORITHM

Genetic algorithm is a promising and upcoming research area for mining association rules. Genetic algorithm is a method which simulates search of evolutionary process. Genetic algorithm can dispose large-scale data gathered in a lot. It is widely applied in mining association rules. Genetic algorithms are typically implemented using computer simulations where optimization is the main criteria for solving the problem. For this problem, members of a space of candidate solutions, called individuals, are represented using abstract representations called chromosomes. The GA consists of an iterative process that evolves a working set of individuals called population toward an objective function, or fitness function. Traditionally, solutions are represented using fixed length strings, especially binary strings, but alternative encodings have also been developed. Genetic algorithm for association rule mining<sup>[8]</sup>

- 1) [Start] Generate random population of  $n$  Chromosomes
- 2) [Fitness] Evaluate the fitness  $f(x)$  of each chromosome  $x$  in the population.
- 3) [New population] Create a new population by repeating the following steps until the new population is complete

- a) [Selection] Select two parent chromosomes from a population according to their fitness.
- b) [Crossover] With a crossover probability cross over the parents to form a new offspring (children).
- c) [Mutation] With a mutation probability mutate new offspring at each locus (position in chromosome).
- d) [Accepting] Place new offspring in a new population.
- 4) [Replace] Use new generated population for a further run of algorithm.
- 5) [Test] If the end condition is satisfied, stop, and return the best solution in current population.
- 6) [Loop] Go to step B

## VIII. CONCLUSION

In this paper, we introduce the multilevel association rule and its two types and its different methods with its advantage. There is also small introduction on fuzzy logic and need of fuzzy logic in association rule and Genetic algorithm. This methods are used for quantitative data. All the algorithm are used in different application. Genetic-Fuzzy association rule mining model is now used at beginning level but it is also used for complex problem in the future.

## ACKNOWLEDGEMENT

I'm extremely obliged to my guide Mr. Lokesh Gagnani devoid of his guidance the work would not have happened and He supported me to solve my difficulties arise & give Valuable suggestions. I would like to pay my sincere gratitude for his endless motivation & support in progress and success of this Survey work.

## REFERENCES

- [1] Abhishek Kajal, Isha Kajajl, "Multilevel Association Rule In Data Mining", IJCSE-Vol3-No3, June-July-2012.
- [2] G.Usha Rani, Vijaya Prakash, Govardhan, "Mining Multi-Level Association Rule Using Pincer Search Algorithm" IJSR-Vol2-No5,May-2013.
- [3] Chun-Hao Chen, Tzung-Pei Hong, Yeong-Chyi Lee, "Multilevel Genetic Fuzzy Mining Algorithm", IEEE, 2011.
- [4] Ehsan Vejdani Mahmoudi, Elahe Sabethin, Masood Niazi Torshiz, Mahrdad Jalali, G Hamarnaz Tadayon Tabrizi, "Multilevel Fuzzy Association Rules Mining Via Determining Minimum Supports And Membership Function", IEEE, 2011
- [5] Yanuang Shen, Jing Shen, Yongjian Fan, "Study On The Application Of Multilevel Association Rule Based On Granular Computing", IEEE, 2010.
- [6] Tzung-Pei Hong, Chu-Tien Chiy, Shyue-Liang Wang, Kuei- Ying Lin, "Mining A Complete Set Of Fuzzy Multiple Level Rules", TFSA-Vol13-No2, June-2011.
- [7] Dr. M. Renuka Devi, Mrs. A. Baby Sarojini, "Application Of Association Rule Mining With Different Database", JGRCS, 2012.
- [8] Indira k, Kanmani s, "Performance Analysis Of Genetic Algorithm For Mining Association Rules", IJCSE, 2012

- [9] Sandeep Kumar Singh, Mr.Ganesh Wayal, Mr.Niresh Sharma,"A Review: Data Mining with Fuzzy Association Rule Mining", IJERT,2012.
- [10] V. Venkata Ramana, M V Rathnamma, A. Rama Mohan Reddy, "Methods for Mining Cross Level Association Rule In Taxonomy Data Structures",IJCA, 2010.
- [11] E.Ramaraj, K.Rameshkumar, "Ranking Mined Association Rule : A New Measure",JTES, 2009.
- [12] K. Rubia, S. Sasikala, "A Comparative Study Of Fuzzy Logic And Weighted Association Rule Mining On Frequent Datasets" , International Journal Of Computer Science & Engineering Technology (Ijcset) Sep. 2014

