DOI: http://dx.doi.org/10.26483/ijarcs.v8i8.4793

ISSN No. 0976-5697

Volume 8, No. 8,September-October 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

MULTILEVEL ASSOCIATION RULE MINING FOR LARGE DATASETS: A REVIEW

Minal Vanarse
Student, CSE Department,
JNEC,Aurangabad, India

Smita Kasar
Assistant Professor,CSE Department,
JNEC,Aurangabad,India

Abstract - Association rule mining is an imperative research issue in domain of data mining,but association rules mining at single concept level lead to uninteresting rules. For large data applications, it is hard to discover solid association rules among data elements at single abstraction level, because of the lack of data in multidimensional space. So finding association rules at multiple abstraction levels leads to knowledge discovery. The discovery of association rules at multiple levels is helpful in numerous applications. Priorwork in field of data mining has yielded proficient techniques for finding multilevel rules. This study aims to review the multilevel association rule mining and different techniques used for mining multilevel association rules from large datasets.

Keywords: Multilevel Association; Rule Mining; Apriori; Genetic Algorithm; Particle Swarm Optimization.

I. INTRODUCTION

Data mining consist of abstraction of knowledge from data. It discovers new patterns and relations in large datasets. Data mining allows user to analyze the data with different dimensions, categorizing the information and summarize the knowledge from this data. The objective of mining information from data is to mine knowledge from large datasets and transforms it into human comprehensible form [1].

Association rule mining has turned out to be both an imperative and generally utilized data mining method. It is used to extract frequent data items, associations and the correlation between data items from datasets. With the wide utilization of PCs and mechanized information gathering tools huge amount of transactional data have been gathered and deposited in databases. These large data are used in business management, communications, finance, marketing, decision support system, etc. [2]. Existing work in association rules mining research focused on mining rules at single concept level, but to mine strong associations from huge amount of data there is need to focus on mining association rules at multiple levels of hierarchy which leads to more specific and concrete information. The main challenge of data mining is to develop fast and efficient algorithm which can handle large data efficiently [4].

This study reviews the Apriori, Genetic Algorithm and Particle Swarm optimization techniques for association rule mining at multiple levels of abstractionfrom large datasets.

II. MULTILEVEL ASSOCIATION RULES

Mining association rulesat multiple levelsfindinteresting relations among data items. Theproblem of multilevel association rules can be described as category tree. The items in dataset are defined as catalogue treeas given in fig. 1 which is a catalogue tree for food mall. Let category such as ‘Dairy’ represent the first level of category and second level is for type i.e. ‘Milk’, and third level represent brand i.e. ‘Amul’.

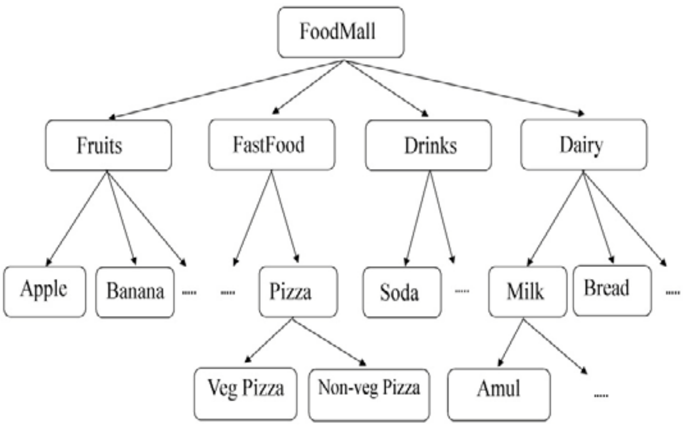

Fig 1: Food mall category tree

A database consists of a set of transactions. Each transaction containsitems in database. Every transaction has unique identifier. The leaves in category tree are items in transaction.

The two imperative measures are used by the association rules are, support and confidence. The support s of ruleindicates how frequently items in rule occur together. The confidence c of rule indicates that probability of both antecedent and consequent appearing in the same transaction.

III. LITERATURE SURVEY

Existing studies on multilevel association rule mining from large datasets involve number of techniques. This section presents a review of some methodologies used to mine the multilevel association rules.

A. *Apriori*

Varioustechniques for association rule mining have been suggested, most of the techniques follow the method proposed by authors in [5] known as Apriori algorithm. It discovers all important association rules amongdata elements in database of transactions. In Apriori candidate itemset is generated, itemset found huge in preceding pass, is only counted without all the transactions in the dataset. The primary perception is that any subset of large dataset must be large. So, itemset having k elements can be produced by linkingitemsets having k-1 elements, and removing the subsets that are not large. In Apriori algorithm transactionsare matched with the candidate itemsets, to decide whether transaction holds the item set or not, and according to its frequent itemsets are generated.

To enhance the performance and efficiency, the variant AprioriTidof Apriori algorithm, is presented in [5] In AprioriTid, in place of the transaction aftereachiteration, an encryption of all the large itemsets in a transaction is used. However in AprioriTid, the transaction is not considered in succeeding iterations, if transaction does not cover any large itemset in present iteration. In [2] a hybrid Apriori algorithm is presented which combines the Apriori and AprioriTid. It uses Apriori for initial process and for later passes it switches to AprioriTid [2].

Another branch for mining association rules is FP-Growth. Authors in [6] described FP growth as it is a divide and conquer strategy. It uses pattern fragment growth method to mine the set of frequent patterns. In FP-growth to avoid costly database scans, large database is compacted into smaller structure. It avoids generation of large number of candidate sets. As compared to Apriori FP-growth is faster, because it requiresa smaller amount of time to search the database and find frequent patterns in multilevel form.

To brought flexibility in association rule mining authors in [8] used OLAP technique with data mining formining association rules at multiple levels. To improve the proficiency of mining multiple level association rules authors in [7] make use of Ant Colony Systems algorithm to mine the multilevel association rules, minimum support is not specified, the support value is determined by the algorithm itself. Authors in [9] optimized the method proposed in [7], in this approach the minimum support is calculated for every item In [10] authors proposed application of fuzzy concept hierarchy for miningassociation rules at multiple levels from large datasets using Attribute-Oriented Induction approach.

In [4] authors proposed a method to mine level crossing association rules from large databases. Mining level crossing rules lead to mine solid associationsbetween items at different levels of abstraction. Authors in [12] presented an effectivetechnique that produces all important relationships between items in large datasets the technique integrates buffer management and cropping techniques. Authors in [11] proposed top-down developingtechnique for mining the multiple-level association rules. Authorsdefine a collection of algorithms 'ML_T1LA', 'ML_T2L1', 'ML_TML1', and'ML_T2LA' based on ways of sharing in-between results. These methods can be developed from large transactional databases for findingstimulating and robust multilevel association rules.

B. Genetic Algorithm

Genetic algorithm is a search and optimization procedure, stimulated by principles of natural selection and natural genetics. It is an adaptive and heuristic search algorithm. For solving problem GA(Genetic Algorithm) is a part of evolutionary computing and it is used to resolve optimization problems, GA mimic the survival of the fittest amongstentitiesoverconsecutive generations. The key inspiration for utilizing Genetic algorithm in mining of association rules is that they implement comprehensive search and handle better with attribute interface than greedy rule induction algorithms used in data mining [1]. In multilevel association rule mining Genetic algorithm can limitassociation rule search space and reach to optimum resultthroughout association rule discovery. Usually for association rule mining all candidates item set has to produce and check them against entire dataset. But by using GA this problem is solved by checking most likely candidates only [3].

To reduce computational cost and heavy database scan Association rule mining with genetic algorithm based methods have been discovered by many researchers. In [13] authors proposed a method which enhance the association rules mined Apriori algorithm, with Genetic Algorithm. By manipulating these rules system can discover the rules comprising negative attributes.

In [14] authors proposed a technique based on Genetic Algorithm in which comprehensive FP-tree is used to implement this method byfinding association rules,where minimum support is not specified. Authors in [15] present a Genetic Algorithm for the prioritizingassociation rules, confidence and strength of the rule; collectively these two measures are used to calculate the fitness function, apart from the support and confidence of mined rules.

In [16] Measures like accuracy, support, and interestingness are used for assessing the rule. Authors proposed a novel method for multi-objective association rule mining using Genetic Algorithms. These measures are used as objective of association rule mining. Authors used Pareto based genetic algorithm for miningvaluable and stimulating rules from transactional database. Authors in [17] also present a Pareto based multi objective progressive algorithm rule mining method based on Genetic Algorithms. To mine the simulating and strong rules from transactional datasets together with the crossover and mutation operator, elitism operator is used. This method does not trustworthy for large transactional dataset.

C. Particle Swarm Optimization

PSO (Particle Swarm Optimization) is a heuristic optimization technique, Kennedy and Eberhart proposed Particle Swarm Optimization in 1995. PSO is Particle swarm optimization is a computational approach that improves an issue by iteratively attempting to develop a candidate solution to a given amount of standards. PSO is motivated by the social behavior of bird flocking or fish schooling. To improve the computational efficiency PSO is used widely in association rule mining [24].

Getting limited relevance rules is important in association rule mining. Therefore for improving the quality of mined rules, to find quality rules authors in [19] proposed a Particle swarm optimization technique. Authors presented a new approach for determining threshold value by algorithm itself, because defining support and confidence is difficult issue in PSO. In this approach the rules are mined using binary encoded data and fitness function. In [20] author proposed a Quantum behaved particle swarm optimization method formining the qualitative association rules in average time for largetransactional datasets with QSO.

Authors in [21] proposed an Artificial Bee Colony Optimization algorithm for hiding sensitive association rules. They used Equivalence Class Transformation (ECLAT) algorithm for finding frequent item-sets using minimum support and minimum confidence measures. For modifying sensitive items, frequent item-sets sensitive data are selected and then Artificial Bee Colony Optimization algorithm is used. Authors in [23] also proposed Artificial Bee Colony Algorithm with one additional operator called crossover for enhancing association rule quality. The crossover operator is used for better exploration ability, as this will generate more number of candidate solutions.

Authors in [18] proposed a Weighed Quantum behaved Particle Swarm Optimization (WQPSO) algorithm for refining the performance of mining association rules. The algorithm determines suitable threshold values by itself and enhances the computational proficiency of Apriori algorithm. In [22] authors proposed a binary particle swarm optimization based association rule mining technique. It generates association rules without stating support and confidence measures. To measure the quality of the rule a fitness function is defined, and product of support and confidence is taken to calculate the fitness function.

IV.CONCLUSION

Mining association rules at multiple level of hierarchy lead to mining of advanced knowledge from large transactional datasets. In this survey the algorithmic aspects of multilevel association rule mining algorithms are studied. The algorithms which are efficient in association rule mining atmultiple levels of abstractions from large transactional datasets are discussed and reviewed.

Algorithms and techniques used for mining association rules at multilevel, generally suffer from high computational cost and efficiency problem. To minimize this, efficient optimization algorithms are required; also there is a requirement to avoid exhaustive scan and database in order to reduce the computational time.

Apriori, Genetic algorithm and Particle Swarm Optimization algorithms were reviewed in the literature. Apriori algorithm computes the frequent itemset exactly, but it goes out of memory or time. Heuristic optimization algorithms like Genetic Algorithm and Particle Swarm Optimization willdecrease the multiple levelassociation rule search space,which results in computational cost reduction. Depending on the application, there is a compromise between efficiency and computational cost of all algorithms.

V. REFERENCES

[1] Freitas, Alex A. 2003, a survey of evolutionary algorithms for data mining and knowledge discovery. Advances in evolutionary computing. Springer Berlin Heidelberg, 819-845.

[2] Savasere, Ashok, Edward Robert Omiecinski, and Shamkant B. Navathe. 1995, an efficient algorithm for mining association rules in large databases. Georgia Institute of Technology.

[3] Xu, Yang, et al. 2014, A genetic algorithm based multilevel association rules mining for big datasets. Mathematical Problems in Engineering.

[4] Thakur, R. S., R. C. Jain, and K. R. Pardasani.2006, Mining level-crossing association rules from large databases. Journal of Computer Science 2.1.

[5] Agrawal, Rakesh, and Ramakrishnan Srikant.1994 ,Fast algorithms for mining association rules. Proc. 20th int. conf. very large data bases, VLDB. Vol. 1215.

[6] Han, Jiawei, Jian Pei, and Yiwen Yin.2000, mining frequent patterns without candidate generation. ACM SigmodRecord.Vol. 29.No. 2. ACM,

[7] Vejdani, E., et al. 2010, Extracting membership functions by ACS algorithm without specifying actual minimum support. Multimedia Computing and Information Technology (MCIT), 2010 International Conference on.IEEE.

[8] Wang, Yingjie, et al. 2013 Improved multi-level association rule in mining algorithm based on a multidimensional data cube. Consumer Electronics, Communications and Networks (CECNet), 2013 3rd International Conference on. IEEE.

[9] Mahmoudi, Ehsan Vejdani, et al. 2011 Multi-level fuzzy association rules mining via determining minimum supports and membership functions. Intelligent Systems, Modelling and Simulation (ISMS), 2011 Second International Conference on.IEEE.

[10] Angryk, Rafal A., and Frederick E. Petry. 2005 Mining multi-level associations with fuzzy hierarchies. Fuzzy Systems, 2005. FUZZ'05. The 14th IEEE International Conference on.IEEE.

[11] Han, Jiawei, and Yongjian Fu. 1995, Discovery of multiple-level association rules from large databases. VLDB.Vol. 95.

[12] Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. Acmsigmodrecord.Vol. 22.No. 2.ACM, 1993.

[13] Saggar, Manish, Ashish K. Agrawal, and Abhimanyu Lad. Optimization of association rule mining using improved genetic algorithms. Systems, Man and Cybernetics, 2004 IEEE International Conference on.Vol.4.IEEE, 2004.

[14] Yan, Xiaowei, Chengqi Zhang, and Shichao Zhang. Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support. Expert Systems with Applications 36.2 (2009): 3066-3076.

[15] Kumar, M. Ramesh, and Dr K. Iyakutti. Application of genetic algorithms for the prioritization of association rules. IJCA special issue on artificial intelligence techniques-novel approaches and practical applications (2011): 1-3.

[16] Ghosh, Ashish, and BhabeshNath. Multi-objective rule mining using genetic algorithms.Information Sciences 163.1 (2004): 123-133.

[17] Wakabi-Waiswa, Peter P., and VenansiusBaryamureeba.Extraction of interesting association rules using genetic algorithms. International Journal of Computing and ICT Research 2.1 (2008): 26-33.

[18] Deepa, S., and M. Kalimuthu. An optimization of association rule mining algorithm using weighted quantum behaved PSO. International Journal of Power Control Signal and Computation 3.1 (2012): 80-85.

[19] Kuo, Ren Jie, Chie Min Chao, and Y. T. Chiu. Application of particle swarm optimization to association rule mining. Applied Soft Computing 11.1 (2011): 326-336.

[20] Ykhlef, Mourad. A quantum swarm evolutionary algorithm for mining association rules in large databases. Journal of King Saud University-Computer and Information Sciences 23.1 (2011): 1-6.

[21] Prabha, M. Sathiya, and S. Vijayarani. Association rule hiding using artificial bee colony algorithm. International Journal of Computer Application (1975-8887) 33.2 (2011).

[22] Sarath, K. N. V. D., and Vadlamani Ravi. Association rule mining using binary particle swarm optimization. Engineering Applications of Artificial Intelligence 26.8 (2013): 1832-1840.

[23] Sharma, Pankaj, Sandeep Tiwari, and Manish Gupta. Association Rules Optimization using Artificial Bee Colony Algorithm with Mutation. International Journal of Computer Applications 116.13 (2015).

[24] Kennedy, J. and Eberhart, R., Particle Swarm Optimization,Proceedings of the IEEE International Conference on NeuralNetworks, Perth, Australia 1995, pp. 1942-1945.

