Multi Objective Fuzzy Association Rule Mining With ABC Algorithm: A Survey

Alok B. Patel[1], Mr. Premkumar Shivakumar [2]

ME Student, Dept. of CE, Silveroak Colleage of Engineering and Technology, Gujarat, India[1]

Assistant professor, Dept. of CE, Silveroak Colleage of Engineering and Technology, Gujarat, India[2]

ABSTRACT

Data Mining is the process of obtaining high level knowledge by automatically discovering information from data in the form of rules and patterns. Data mining is most commonly used in attempts to induce association rules from transaction data. Association rule mining is a well-established method of data mining that identifies significant correlations between items in transactional data. An association rule is an expression XàY, where X and Y are a set of items. It means in the set of transactions. Fuzzy Association rule mining is an essential topic in Information retrieval mining field and produces all important Fuzzy association rules between attributes in the dataset because large data set records considered as transactions. In ABC algorithm, the position of a food source represents a possible solution to the optimization problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. In this paper ABC seems particularly suitable for multi-objective optimization mainly because of solution quality and the high speed of convergence that the algorithm presents for single-objective optimization. Proposed work examines a new Multi-Objective Fuzzy Rule Mining with ABC algorithm.

KEYWORDS: *Data mining; Rule optimization, Artificial Bee Colony; Fuzzy Association rule mining, Multi-Objective*

I.INTRODUCTION

Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. Now a days, Data mining has attracted a great deal of attention in the information industry and in society as a whole due to the wide availability of large amounts of data and the imminent need for turning such data into useful information and knowledge which can be used for applications ranging from

Market analysis, fraud detection, and customer preservation, to production control and science exploration. In some cases, users may have no idea regarding what kinds of patterns in their data may be interesting, and hence may like to discover for several different kinds of patterns in parallel. Thus it is essential to have a data mining system that can mine multiple kinds of patterns to accommodate different user expectations or applications. Furthermore, it should be able to discover patterns at various granularities [1].

Association Rule Mining (ARM) is one of the foremost imperative research areas in the concept of data mining that facilitate the mining of concealed repeated patterns that based on their own frequencies in the shape of association rules from any item set or datasets containing entities to represent the most recent trends in the given dataset. These mined repeated patterns or fuzzy association rules uses either for physical data analysis or additionally influenced to compel any mining tasks like categorization and collecting which helps domain area experts to automate decision-making solutions. Now a day's FARM has deliver a good tremendous recognition owing of its correctness or accurateness, which might be described to its capability to mine massive amounts of knowledge from very large transactional and relational datasets. Currently frequent patterns retain all the prevailing relationships between items and entities in the given dataset and pact only with the numerically noteworthy associations, classification or clustering. Association rules mining technique in widely used in various areas such as telecommunication networks, stock market research and risk management, inventory control etc. The Apriori algorithm is used for frequent item set mining using association rules over the transactional databases. The apriori algorithm is proceeds by recognize the frequent individual items in the dataset and expanding them to larger and larger item sets as long as those item sets appear adequately often in the database[2].

II.CLASSIFICATION OF ASSOCIATION RULE MINING ALGORITHMS

Association rule mining algorithms can be divided in two basic classes; these are BFS like algorithms and DFS like algorithms [2]. In case of BFS, at first the minimum support is determined for all item sets in a specific level depth, but in DFS, it descends the structure recursively through several depth levels. Both of these can be divided further in two sub classes; these are counting and intersecting. Apriori algorithm comes under the counting subclass of BFS class algorithms. It was the first attempt to mine association rules from a large dataset. The algorithm can be used for both, finding frequent patterns and also deriving association rules from them. FP-Growth algorithm falls under the counting subclass of DFS class algorithms. These two algorithms are the popular example of the classical association rule mining.

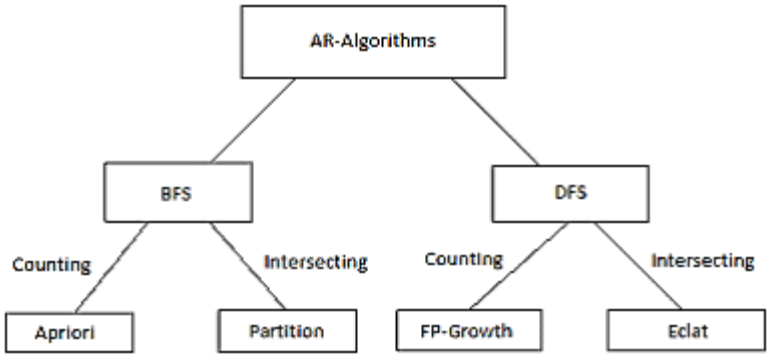

Figure 1: Classification of Mining Algorithm

III. CLASSICAL ASSOCIATION RULE MINING AND FUZZY ASSOCIATION RULE MINING

Classical association rule mining depends on the Boolean logic to transform numerical attributes into Boolean attributes by sharp partitioning of dataset. So that number of rules generated is low. It is inefficient in case of huge mining problems. In the classical association rule mining algorithms users have to specify the minimum support for the given dataset on which the association rule mining algorithm will be apply. But it is very much possible that the user sets a wrong minimum support value which can hamper the generation of association rules. And the setting of minimum support is not an easy task. If minimum support is set to a wrong value then there is a big possibility of combinatorial blow up of huge number of association rule within which many association rules will not be interesting. Fuzzy association rule mining first began in the form of knowledge discovery in Fuzzy expert systems. Instead of Boolean logic, a fuzzy expert system [3] uses a collection of fuzzy membership functions and rules [4]. The rules in a fuzzy expert system are usually of a form similar to the following: "If it is raining then put up your umbrella"

Here if part is the antecedent part and then part is the consequent part [5]. This type of rules as a set helps in pointing towards any solution with in the solution set. But in case of Boolean logic every data attribute is measured only in terms of yes or no, in other words positive or negative. So it never allows us to have the diverse field of solutions. It always marginalizes the solutions; on the other hand fuzzy logic keeps broad ways of solutions open for the users. Many other fuzzy logic techniques are also use in fuzzy association rule mining [6]. Classical association rule mining uses the concept of crisp sets.

IV. FUZZY ASSOCIATION RULE MINING ALGORITHMS

In the last few decades there has been a large number of research work already done in the field of fuzzy association rule mining. The concept of fuzzy association rule mining approach generated from the necessity to efficiently mine quantitative data frequently present in databases. Algorithms for mining quantitative association rules have already been proposed in classical association rule mining. Dividing an attribute of data into sets covering certain ranges of values, engages the sharp boundary problem. To overcome this problem fuzzy logic has been introduced in association rule mining. But fuzzy association rule mining also have some problems.. Classical association rule mining regarding the sharp partitioning. Following are some partitioning rule:

• Use of sharp ranges creates the problem of uncertainty. More precisely loss of information happens at the boundaries of these ranges. Even at the small changes in determining these intervals may create very unfamiliar results which could be also wrong.

• These partitions do not have proper semantics attached with them.

In fuzzy association rule mining the transformation of numerical attributes into fuzzy attributes is done using the fuzzy logic concept. In fuzzy logic attribute values are not represented by just 0 or 1. Here attribute values are represented within a range between 0 and 1[7]. According to this way, crisp binary attributes are converted to fuzzy attributes and by using fuzzy logic; we can easily resolve the above problems. The algorithms which are mostly use for fuzzy association rule mining are the fuzzy versions of Apriori algorithm. Apriori algorithm is slow and inefficient in case of large datasets. Fuzzy versions of Apriori algorithm would not be able to handle real-life huge datasets. Algorithms uses the principle of memory dependency like FP-Growth and its fuzzy versions are inadequate to deal with huge datasets. But these huge data sets can be easily managed by the partial memory dependent variant algorithms like ARMOR and. AshishMangalampalli, VikramPudi [8] proposed a new fuzzy association rule mining algorithm which will perform mining task on huge datasets efficiently and in fast. Their proposed algorithm has two-step processing of dataset. But before the actual algorithm there is preprocessing of dataset by fuzzy c-means clustering. Fuzzy partitions can be done on given data set so that every data point is a member of each and every cluster with a certain membership value. Main objective of the algorithm is to minimize the Equ (1)2 (1) Where m is any real number such that $1 \leq m < \infty$, μij is the degree of membership of xiin the cluster of j, xi is theith dimensional measured data,$cj$is the d-dimensional cluster center, and is any norm expressing the similarity between any measured data and the center. By this way corresponding fuzzy partitions of the dataset is generated where each

value of numeric attributes are uniquely identified by their membership functions (μ). Depending upon the number of fuzzy partitions defined for an attribute, each and every existing crisp data is converted to multiple fuzzy data. This has the possibility of combinatorial explosion of generation of fuzzy records. So they have set a low threshold value for the membership function μ which is 0.1 to keep control over the generation of fuzzy records. During the fuzzy association rule mining process, the original data set is extended with attribute values within the range (0, 1) due to the large number of fuzzy partitions are being done on each of the quantitative attribute. To process this extended fuzzy dataset, some measures are needed which are based on t-norms [8], [9], [10]. In this way the fuzzy dataset E is created upon which the proposed algorithm will work. The dataset is logically divided into p disjoint horizontal partitions*P*1, 2, *P*. Each partition is as large as it can fit in the available main memory.

V. EVOLUTIONARY MULTIOBJECTIVE CLASSIFICATION RULE MINING

Evolutionary multi objective techniques in classification rule mining can be roughly categorized into two approaches. In one approach, each rule is evaluated according to multiple rule evaluation criteria such as support and confidence. An EMO algorithm is used to search for Pareto-optimal classification rules. In the other approach, each rule set is evaluated according to multiple rule set evaluation criteria such as accuracy and complexity. An EMO algorithm is used to search for Pareto-optimal rule sets. In this section, we briefly explain these two approaches.

VI. LITERATURE REVIEW

An enough literature is available for multi objective fuzzy association rule mining and also classifies using various approaches.

K.Satheshkumar [11] An Hybrid Optimization Algorithm for Fuzzy Association rule Mining discussed Fuzzy association rule mining, considerably minimizes the effectiveness of the rules and turns it difficult for the user to use and decide on these rules. It produces all important Fuzzy association rules between attributes in the dataset because large data set records considered as transactions. Each Transaction consists of set of attributes. This is attributed to generation of more number of rules. All the generated rules do not produce best classification results and also it takes more number of times. To overcome this problem we proposed Fuzzy based artificial bee colony optimization algorithm which picks the best rules from the given population of rules.

Diana Mart´ın[12]**A New Multi objective Evolutionary Algorithm for Mining a Reduced Set of Interesting Positive and Negative Quantitative Association Rules** we propose MOPNAR, a new multi objective evolutionary algorithm, in order to mine a reduced set of positive and negative quantitative association rules with low computational cost. To accomplish this, our proposal extends a recent multi-objective evolutionary algorithm based on decomposition to perform an evolutionary learning of the intervals of the attributes and a condition selection for each rule, while introducing an external population and a restarting process to store all the no dominated rules found and to improve the diversity of the rule set obtained. Moreover, this proposal maximizes three objectives—comprehensibility, interestingness, and performance—in order to obtain rules that are interesting, easy to understand, and provide good coverage of the dataset.

HuiZheng[13]**Optimized Fuzzy Association Rule Mining for Quantitative Data** an optimized fuzzy-association-rule mining algorithm based on a generic measure has been proposed. We have shown that the features of the multiple objective function optimizations make the proposed model easy to formulate and use for continuous data.

Taking the two-level iteration processes into account, the fuzzy association rules and the frequent item-sets are optimized by improving the fuzzy-set partition parameters repeatedly.

Yanjiao Wang[14]**Multi-objective Artificial Bee Colony algorithm** multi-objective algorithm based on artificial bee colony algorithm has been proposed in this paper. New adaptive searching operations are designed for employed bees and onlookers, which makes it not difficult to determine the better individuals in MOPs. Meanwhile, the approach of determining elite population is improved. The results on several general test functions have shown that solutions obtained by MABC algorithm are closer to the true Pareto no dominated front and distribute more uniformly than other multi objective optimization algorithms.

Chun-Hao Chen[15]**MOGA for Multi-Level Fuzzy Data Mining**a Multi-Objective MultiLevel Genetic-Fuzzy Mining (MOMLGFM) algorithm for mining a set of non-dominated membership functions for mining multilevel fuzzy association rules. a MOGA fuzz data mining approach for deriving sets of membership functions suitable for multiple-level association rule mining According to the given taxonomy, the proposed algorithm first encodes the set of membership functions of item classes (categories) into a chromosome with real-number schema. In the evolution process, the two objective functions are used for evaluating individuals. Experiments on a simulation dataset were also made to show the effectiveness of the proposed approach.

VI.CONCLUSION

Based on our survey Knowledge extraction in databases may be the method of extracting data within the type of interesting rules. These rules are domain specific. These rules reveal the association relationship among totally different data's that however a specific information items expounded to a different information item. So, we have a tendency to decision these rules as association rule. These rules are heuristic in nature. The method of extracting and managing these rules is understood as association rule mining. Association rule mining is a very important method in intelligent systems like Expert system. As a result of these intelligent systems solves domain specific issues. This will even cause the generation of large number of redundant rules further more as useless rules. So, it is very a difficult task of setting an accurate minimum support value manually. That is why classical association rule mining is time consuming and fewer accurate methods. Fuzzy association rule mining is comparatively a more recent idea. This uses the idea of fuzzy set theory for mining job.

REFERENCES

[1] J. Han and M. Kamber, "Data mining: Concepts and techniques (2nd edition)," Morgan Kaufman Publishes, 2006.
[2] Hipp, Jochen; Guentzer, Ulrich; Nakhaeizadeh, Gholamreza: Algorithms for Association Rule Mining - A General Survey and Comparison. ACM SIGKDD Explorations Newsletter, Volume 2, Issue 1, 2000.
[3] Türks¸en, I.B. and Tian Y. 1993. Combination of rules and their consequences in fuzzy expert systems, Fuzzy Sets and Systems, No. 58,3-40, 1993.
[4] http://www.cs.cmu.edu/Groups/AI/html/faqs/ai/fuzzy/ part1/faq-doc-4.html.
[5] L.A.Zadeh.Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on System, Man, and Cybernetics, Volume 3, Pages(s):28-44, January, 1973.
[6] Delgado, Miguel: Fuzzy Association Rules: an Overview. BISC Conference, 2003.

[7] AshishMangalampalli, VikramPudi: Fuzzy Association Rule Mining Algorithm for Fast and Efficient Performance on Very Large Datasets. FUZZ-IEEE 2009, Korea, ISSN: 1098-7584, E-ISBN: 978-1- 4244-3597-5, Page(s): 1163 – 1168, August 20-24, 2009.
[8] R.R. Yager, "On Linguistic Summaries of Data," in, pp. 347-363.
[9] Yan, P., Chen, G., Cornelis, C., De Cock, M., Kerre, E.E.: Mining Positive and Negative Fuzzy Association Rules. In: KES, Springer, Page(s): 270-276, 2004.
[10] Verlinde, H., De Cock, M., Boute, R.: Fuzzy Versus Quantitative Association Rules: A Fair Data-Driven Comparison. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, Volume 36, Page(s): 679-683, 2006.
[11] K.satheshkumar, M. Hemlatha, "An Hybrid Optimization Algorithm for FuzzyAssociation rule Mining," IEEE 2012.
[12] Diana Mart´ın, Alejandro Rosete, Jes´usAlcal´a-Fdez, Francisco Herrera, "A New Multiobjective Evolutionary Algorithm for Mining a Reduced Set of Interesting Positive and Negative Quantitative Association Rules," IEEE 2014.
[13] HuiZheng, Jing He, Guangyan Huang, Yanchun Zhang,"Optimized Fuzzy Association Rule Mining for Quantitative Data," IEEE 2014.
[14] Yanjiao Wang, Yaojie Li, "Multi-objective Artificial Bee Colony algorithm," IEEE 2015.
[15] Chun-Hao Chen, Chi-HsuanHo, Tzung-Pei Hong, Wei-Tee Lin, "MOGA for Multi-Level Fuzzy Data Mining," IEEE 2012.
[16]