Fuzzy Association Rule Mining and ABC Algorithm: A Survey

Alok B Patel[1], Apexa J Bhavsar[2], Aditi M Joshi[3]
[1, 2, 3] Assistant professor, Dept of CE
[1, 2, 3] Swarrnim Start up and Innovation University, Gujarat, India

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 algorithms are inspired by some natural phenomenon and called Nature Inspired Algorithms (NIAs). The NIAs mimics the intelligent behavior of social insects like bees, ants, termites, fish, birds, etc. Swarm Intelligence getting popularity now days and become a rising and fascinating area. It depends on the cooperative behavior of societal living thing. Societal individual makes use of their skill of societal wisdom to crack multifaceted everyday jobs. The main power of swarm based optimization strategy is multiple interactions in societal colonies. Swarm intelligence strategies have the potential to solve complex factual world optimization problems as the preceding study have exposed. This survey provides a of ABC algorithm and analysis of its performance in various sector.*

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]. 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.

Swarm Intelligence based algorithm are very popular now a days while solving complex problem in field of engineering, management and science. These algorithms are inspired by some natural phenomenon and called Nature Inspired Algorithms (NIAs). The NIAs mimics the intelligent

behaviour of social insects like bees, ants, termites, fish, birds, etc. Swarm Intelligence getting popularity now days and become a rising and fascinating area. It depends on the cooperative behaviour of societal living thing. Societal individual makes use of their skill of societal wisdom to crack multifaceted everyday jobs. The main power of swarm based optimization strategy is multiple interactions in societal colonies. Swarm intelligence strategies have the potential to solve complex factual world optimization problems as the preceding study have exposed.

The ABC strategy is somewhat a clear-cut, swift and population based stochastic search strategy in the area of nature-inspired optimization strategies. The position of swarm updates in ABC by two contradictory activities: first one is a process of adaptation, which empowers exploring the diverse search space, and the second one is a process of selection, which ensures the exploitation of the earlier experience. Sometimes it is observed that ABC stops moving in the direction of global optimum despite the fact that the population has not congregate to a restricted most advantageous [6]. It can be experiential to facilitate the solution investigation equation of ABC is fine at exploration, however pitiable at exploitation [3]. For that reason, it is enormously enviable to develop a new approach which is able to exploit better solutions in its neighborhood and also able to explore the search space for less fit solutions in order to uphold the appropriate balance among exploration and exploitation activities of ABC.

The ABC met heuristic technique is stimulated through the spontaneous food foraging behavior of the honey bee creature. Honey bee insect most intuitive creation of nature; it shows combined intellectual behavior at the same time as penetrating the food. The honey bee can memorize the ecological circumstances, can accumulate and distribute the information and can decide according to these observations. As per the changes in the surroundings, the bee updates its position, assign the responsibilities dynamically and go on further by means of societal erudition and education. This extraordinary conduct of honey bees motivates research scientists to imitate the intellectual food foraging behavior of the bees.

II. CLASSICAL ASSOCIATION RULE MINING AND FUZZY ASSOCIATION RULEMINING

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 crispsets.

III. FUZZY ASSOCIATION RULE MININGALGORITHMS

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. Ashish Mangalampalli, Vikram Pudi [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. 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$P1, 2, P$. Each partition is as large as it can fit in the available main memory.

IV. LITERATUREREVIEW

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

Lijun Sun [11] **An Artificial Bee Colony Algorithm with Random Location Updating paper presents an artificial bee colony algorithm with random location updating,** and the search equation of this algorithm takes a random position in swarm population as the search center. In contrast to ABC, the search range of new solution is further expanded, which can enhance the exploration ability. Besides this, the chaos is used to initialize the swarm population, and diversity of initial population is improved. Then the tournament selection strategy is adopted to maintain the population diversity in the evolutionary process. The results of the simulation experiment on a suite of unconstrained benchmark functions demonstrate that RABC not only has stronger exploration ability but also has better effect on convergence speed and optimization precision, and it can keep good robustness and validity with the increase of dimension.

Sangeeta Sharma [12]**Artificial Bee Colony Algorithm: A Survey** of science where accuracy and efficiency are major issue while solving complex real world problem from almost each and every field of life demands for robust, fast, and accurate optimizers. The ABC algorithm is one of the simplest swarm intelligence algorithms that offer good result with accuracy for optimization problems with different level of complexity. The ABC algorithm proves that it is best choice when tested for standard benchmark problems and complex real world problem. It can be applied for large class of problems. The ABC algorithm is not insightful to primary parameter values and not affected by the increasing dimension of the problem. The ABC algorithm has some drawbacks also, as sometimes it suffers from premature convergence or stagnation that results in loss of balance between intensification and diversification capabilities. This paper endeavors to present a state of art survey of ABC algorithm.

Harathi Parasur Babu[13] **Spatial data mining using association rules and fuzzy logicfor autonomous exploration of geo-reference in data mining** strategy using association rules of a general nature can be applied in biological databases for understanding the hidden information about the disease and its causative agents. Interest among different user groups (physicians, researchers, health care administrators, and policy makers) creates expectations regarding the results and active role in the development and in interactive use of the information. Several design cases in which data mining could be implemented to enable efficient and user friendly knowledge extraction (Kanza and Babic 2014). Fuzzy logic is a method to render precise what is imprecise in the world of medicine where the conversion of the risk to quantitative information may be of considerable importance in counseling patients. By accurately delineating risk, patients will be able to concretely identify risk factors

and ascertain the impact of lifestyle changes thus providing positive reinforcement. Furthermore, this may serve as the basis for future paradigms that determine individualized cancer prevention strategies in subjects harboring an inherited risk.

Ulas Cinar1 [14]**A Hybrid Risk Assessment Method for Mining Sector Based on QFD, Fuzzy Logic and AHP**, The widely used methods for risk analysis and risk assessment in the mining sector are based on the foundation of "probability x intensity". It makes possible to provide assessments using lateral additions and different perspectives. Furthermore, the utilizing data for these methods are generally in categorical structure. The major disadvantage of these methods is that it is possible to obtain very different results when the same assessment is made by different occupational safety specialists. The main reason of this is that occupational safety specialists have different expertise knowledge and they use subjective categorical data. For the assessment, the degree of expertise and categorical data includes both imprecise and vagueness information. For this reason, fuzzy set theory is preferred to cope with imprecise and vagueness structure in a problem. The second problem for the risk assessment is that none of the current method consider the interactions among the activities occurred at work site. All the risk assessment methods consider the activities as independent events. However, all activities have interactions and interdependencies among them, and they provide butterfly effect. Therefore, a new risk assessment approach has been proposed to literature to overcome the deficiencies given above.

Chun-Hao Chen[15]**MOGA for Multi-Level Fuzzy Data Mining** a Multi-Objective Multi Level 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.

V. 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.HanandM.Kamber,―Datamining:Conceptsandtechniques(2ndedition),‖MorganKaufmanPublishes, 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,―OnLinguisticSummariesofData,‖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] Lijun Sun,Tianfei Chenan and Qiuwen Zhang 2An Artificial Bee Colony Algorithm with Random Location

Updating-Hindawi Scientific Programming Volume 2018, Article ID 2767546, 9 pages https://doi.org/10.1155/2018/2767546

[12] Sangeeta Sharma and Pawan BhambuArtificial Bee Colony Algorithm: A Survey,International Journal of Computer Applications (0975 – 8887)

[13] Spatial data mining using association rules and fuzzy logicfor autonomous exploration of geo-referenced cancer datain Western Tamilnadu, India-Received: 12 June 2014 / Revised: 30 June 2015 / Accepted: 22 July 2015Springer-Verlag Wien 2015

[14] A Hybrid Risk Assessment Method for Mining Sector Based on QFD, Fuzzy Logic and AHP Ulas Cinar1 and Selcuk Cebi2(&)© Springer Nature Switzerland AG 2020 C. Kahraman et al. (Eds.): INFUS 2019, AISC 1029, pp. 1198–1207, 2020. https://doi.org/10.1007/978-3-030-23756-1_141

[15] HuiZheng,JingHe,GuangyanHuang,YanchunZhang,―OptimizedFuzzyAssociationRuleMiningfor Quantitative Data,‖ IEEE2014.