MULTI OBJECTIVE FUZZY ASSOCIATION RULE MINING WITH ABC ALGORITHM

Alok B. Patel[1], Mr. PremkumarShivakumar[2]
ME Student, Dept. of CE, Silver Oak college of Engineering and Technology, Gujarat, India[1]
Assistant professor, Dept. of CE, Silver Oak College 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 seeks to discover knowledge that is accurate, comprehensible and interesting .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 X Y, where X and Y are a set of items. It means in the set of transactions. If all the items in X exist in a transaction. Then Y is also in the transaction with a high probability. 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. Each transaction consists of set of attributes. Multi-objective is an area of multiple criteria decision making, that is concerned with mathematical optimization problems involving more than one objective function to be optimized simultaneously. A MOO problem with constraints will have many solutions in the feasible region.**
Even though we may not be able to assign numerical relative importance to the multiple objectives, we can still classify some possible solutions as better than others. 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. On inspecting the behavior of real bees on finding nectar and sharing the information of food sources to the bees in the hive. 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.

***Index Terms*—Data mining; Rule optimization, Artificial Bee Colony; Fuzzy Association rule mining, Multi-Objective**

I. INTRODUCTION

The rapid advancement in optimization studies has vitally achieved non-optimized design of algorithm (s) with intended or desired functionalities in recent years. Many analytical methods do not provide the feasible solution, which can be obtained using optimization algorithm within reasonable time consumption [1]. The ability to give optimal solution to every optimization run enables selection of feasible solutions to provide the criteria. Most of the real life problems can be optimized in major areas on computer science and mathematics. Optimization can be categorized into four types, combinational optimization, evolutionary algorithm, gradient methods and stochastic optimization. Most importantly, population based search engine algorithm used in operational research and computer science is Bee Colony optimization. This kind of algorithm prepares the search by neighborhood search and random search or it can use both the types of search optimization. Swam Intelligence fields, have adopted Bee colony optimization techniques. The agents are artificial bees which solves complex optimization problems. This behavior of swam characteristics with social insects, communication between the colony insects representing Artificial Bee Colony(ABC) algorithm is swam intelligence for solving combinational optimization problems, was proposed by Karaboga in 2005 [1]. It involves foraging behavior of the bee algorithm for solving rule optimization problems and mainly contains three groups, first one employee bee, second on lookers and thirdly, scouts. The first half of the colony consist of the employee bees and second half consist of on looker bees for every food source, there is only one employee bee, in other words the number of employee bee is equal to food source, the employee bee of an abounded food source become scouts [1].

Association, data mining technique among the various data mining methods, rules to find the necessary patterns from the raw set of patterns facilitating categorize the patterns with relations. Necessary association rules are mined from available association rules by most association rule mining technique for decision making. Fuzzy association rule mining uses Fuzzy logic to create and convert numerical attributes to Fuzzy attributes to create a Fuzzy logic and it can be used to convert numerical attributes to Fuzzy attributes. The popular Fuzzy association rule mining algorithm is available today as Fuzzy Apriori and its different variations have well supported for the Fuzzy numerical analysis. The proposed work adapted Fuzzy Apriori algorithm using Fuzzy Artificial bee colony optimization (FABCO).

II. RELATED WORK

One of the most popular research tasks in data mining is the discovery of frequent item sets and association rules. Fuzzy association rules are in generally understandable by humans easily because of the linguistic terms associated with Fuzzy sets. In addition, many researchers have proposed a Fuzzy association rule mining techniques.

Proposed work examines a new multi objective fuzzy algorithm based on a modified Artificial Bees Colony algorithm. However this algorithm still suffers from some drawbacks, such as local optima and sensitivity to initialization [16].

III. PROPOSED WORK

A. Artificial Bee Colony (ABC) algorithm

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. The number of the employed bees or the onlooker bees is equal to the number of solutions in the population. The main steps of the algorithm are highlighted below [17].

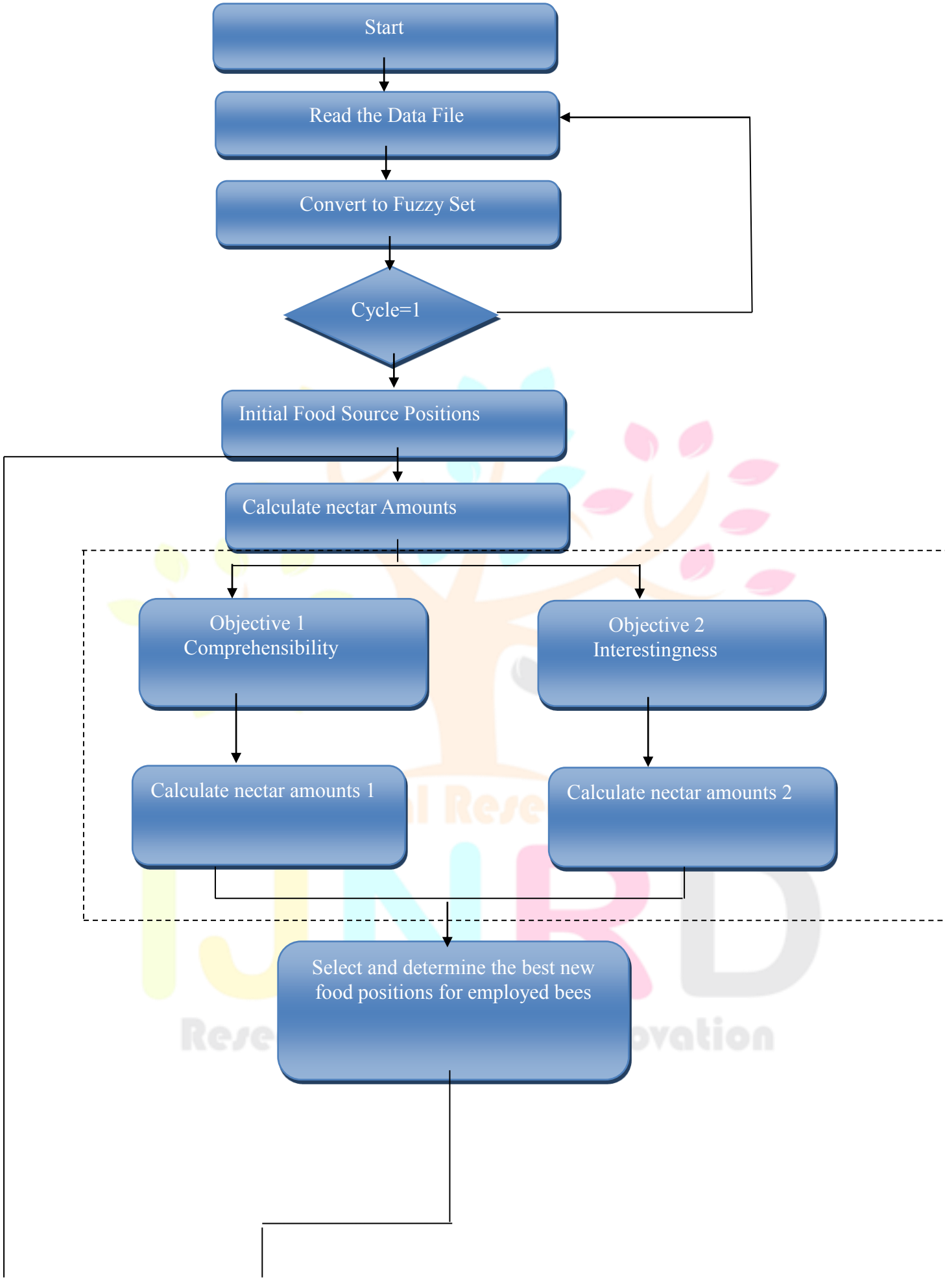
1: Initialize Population
2: repeat
3: Place the employed bees on their food sources

4: Place the onlooker bees on the food sources depending on their nectar amounts
5: Send the scouts to the search area for discovering new food sources
6: Memorize the best food source found so far
7: until requirements are met

Totally, ABC algorithm employs four different selection processes:
(1) A global probabilistic selection process, in which the Probability value is calculated by (2) used by the onlooker bees for discovering promising regions, (2) A local probabilistic Selection process carried out in a region by the employed bees and the onlookers depending on the visual information such As the color, shape and fragrance of the flowers (sources) (bees will not be able to identify the type of nectar source until they Arrive at the right location and discriminate among sources growing there based on their scent) for determining a food source Around the source in the memory as defined (1). (3) A local selection called greedy selection process carried out by Onlooker and employed bees in that, if the nectar amount of the candidate source is better than that of the present one, the bee forgets the present one and memorizes the candidate source produced by (7). Otherwise, the bee keeps the present one in the memory. (4) A random selection process carried out by scouts as defined in (8). It is clear from the above explanation that there are three control parameters in the basic ABC: The number of food sources which is equal to the number of employed or onlooker bees (SN), the value of limit and the maximum cycle number (MCN). In the case of honeybees, the recruitment rate represents a measure of how quickly the bee colony finds and exploits a newly discovered food source. Artificial recruiting could similarly, represent the measurement of the speed with which the Feasible solutions or better quality solutions of the difficult optimization problems can be discovered. The survival and progress of the bee colony are dependent upon the rapid discovery and efficient utilization of the best food resources. Similarly, the successful solution of difficult engineering problems is connected to the relatively fast discovery of good solutions especially for the problems that need to be solved in real time. In a robust search process, exploration and exploitation processes must be carried out together. In the ABC algorithm, while onlookers and employed bees carry out the exploitation process in the search space, the scouts control the exploration process.

B. Proposed Multi Objective Fuzzy Artificial Bee colony optimization (MOFABCO)

The proposed work presents an ABC algorithm for the specified problem to minimize the number of Fuzzy association rules. Fuzzy based Apriori algorithm uses transaction data set which deploys user interested support and confidence value resulting the Fuzzy association rule set. These Fuzzy association rule sets are discrete and continuous hence; need to prune weak rule sets. Optimization of result is needed, henceforth; Multi Objective Fuzzy Artificial Bee colony optimization (MOFABCO) algorithm for Fuzzy association rule was accomplished. Using the fitness value for finding the probability of occurrence of the each rule the greedy selection process will be decided and then optimized by Fuzzy association rule set which is generated. In this proposed work, the adoption of FABCO the resultant rules of Fuzzy association rules are considered as the population. In order to find the best rules, the fitness value of each rule is identified.


Start
Read the Data File
Convert to Fuzzy Set
Cycle=1
Initial Food Source Positions
Calculate nectar Amounts
Objective 1
Comprehensibility
Objective 2
Interestingness
Calculate nectar amounts 1
Calculate nectar amounts 2
Select and determine the best new food positions for employed bees

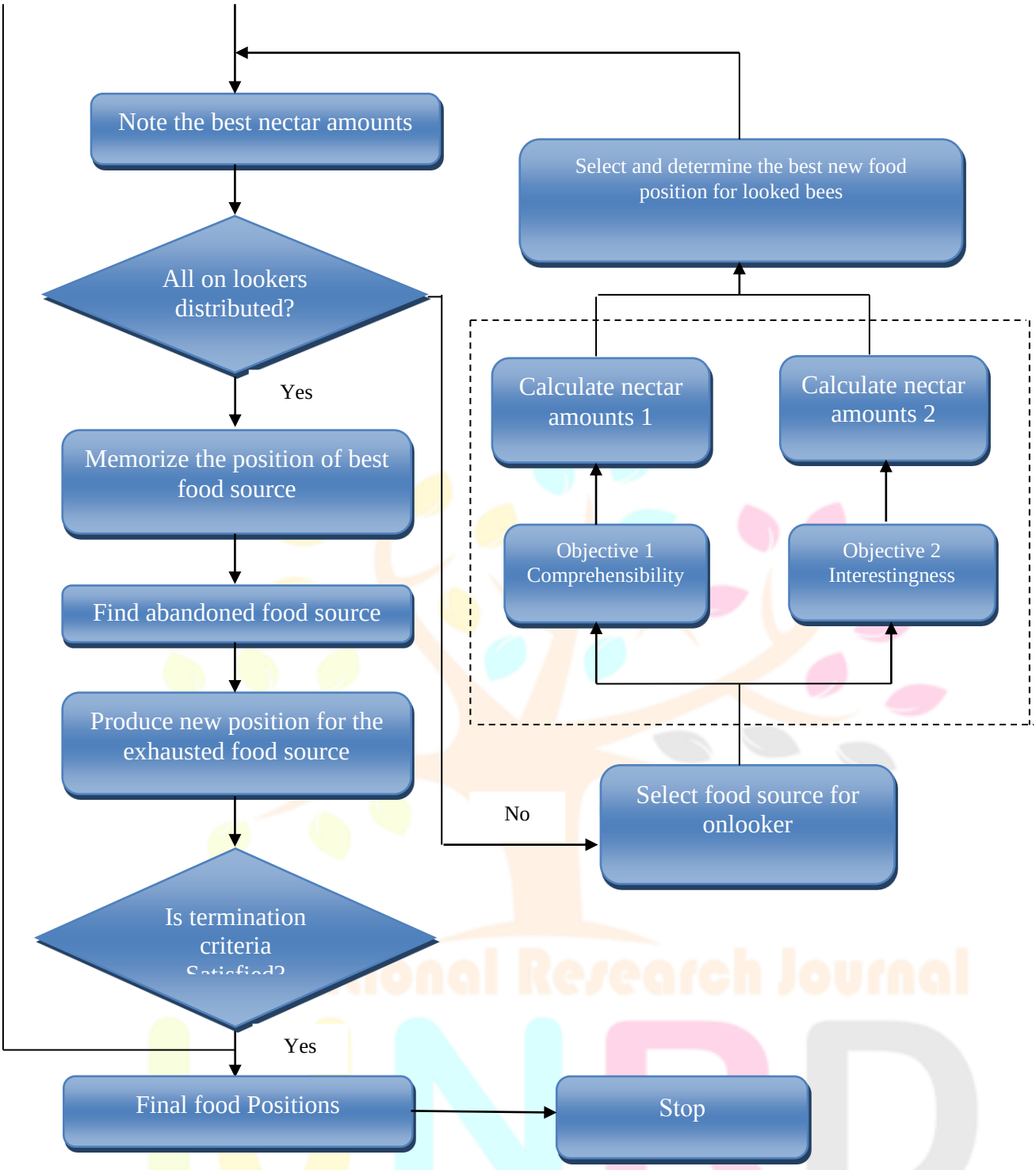


Fig 1: Proposed System

After Producing Vi within the boundaries, a fitness value for minimization problem can be calculated to the solution Vi by

Steps

Step 1: Read the Data File
Step 2: Convert to Fuzzy Set (Low, Medium, High)
Step 3: Initialize the population of solutions X, I = 1… SN
Step 4: Evaluate the population
Step 5: Cycle = 1
Step 6: Return Step 1
Step 7: Produce new solutions ti for the employed bees by using the Multi Objective Functions
  a. Objective 1 (Comprehensibility)
  b. Objective 2 (Interestingness)
Step 8: Apply the greedy selection process for the employed bees
Step 9: Calculate the probability values Pi,j for the rules xi,j using fitness of the solution

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n}$$

Step 10: For each onlooker bee, produce a new solution υi by.

$$v_{ij} = X_{ij} + \emptyset_{ij}(X_{ij} - X_{kj})$$

Step 11: Apply selection process between υi and xi based on greedy method for onlooker bee.

Step 12: If Scout Production Period (SPP) is completed, determine the abandoned solutions by using limit parameter for the scout, if it exists, replace it with a new randomly produced solution by

$$X_i^j = X_{min}^j + \text{rand}\,[0,1]\,(X_{max}^j - X_{min}^j)$$

Step 13: Memorize the Best Solution

Step 14: Cycle = Cycle + 1

Step 15: Until Cycle = MCN

IV. EXPERIMENTAL RESULTS

For measuring the performance of the proposed MOFABCO data mining algorithm, an experimental run with dataset "Indian Pima Diabetes". selected from the UCI Machine Learning Repository was conducted. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of "Pima Indian Diabetes" heritage. An adaptive learning routine that generates and executes digital analogs of perception-like devices.

Attribute Information:

Number of times pregnant

1. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
2. Diastolic blood pressure (mm Hg)
3. Triceps skin fold thickness (mm)
4. 2-Hour serum insulin (mu U/ml)
5. Body mass index (weight in kg/ (height in m) ^2)
6. Diabetes pedigree function
7. Age (years)
8. Class variable (0 or 1)

Fig. 2 shows the Pima Indians Diabetes dataset our proposed work reduces the rules to generated optimized classification result efficiently compare to FABCO technique the accuracy of the proposed work performs better after the pruning process.

Fig. 2 Reduced Rule Details

FABCO	MFABCO
96.46915	96.53699
96.05039	96.58352
96.95342	97.06216
96.57719	96.96017
97.06051	97.06177

96.09223	96.32026
96.55522	96.79409
96.68592	97.10845
96.45089	96.89566
95.58532	96.32424

Table 1 Accuracy Comparison

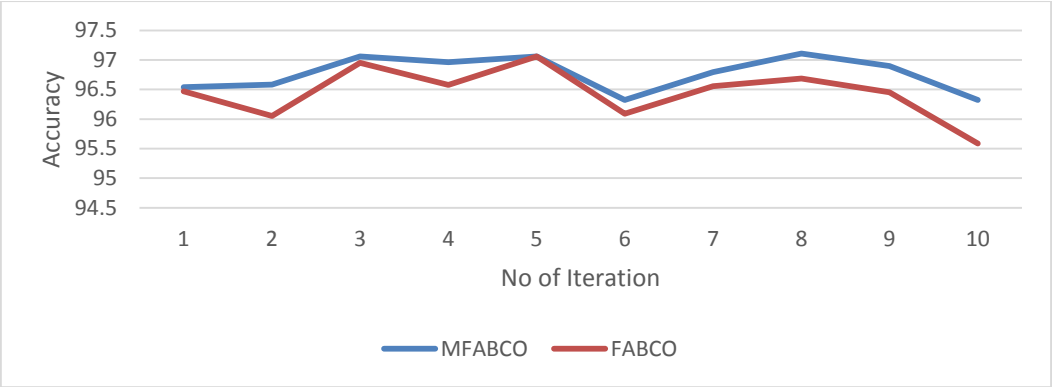


Fig. 3 Graphical Representation of Accuracy

V. CONCLUSION

As we done research, we concluded that Multi Objective Optimization plays very important role for better optimization of Fuzzy Association Rules. We have proposed MFABCO that allows mining with a reduced dataset. A dataset easy to understand, interesting and offer good coverage of the dataset two objective: Comprehensibility and Interestingness. All the generated rules do not produce best classification results and also it takes more number of times. MFABCO algorithm which picks the best rules from the given population of rules.

ACKNOWLEDGEMENT

We thank the Gujarat Technological University and Silver Oak Institute of Engineering for the motivation and encouragement for giving the opportunity to do this research work as successful one.

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 Fuzzy Association rule Mining,‖ IEEE 2012.
[12] Diana Mart´ın, Alejandro Rosete, Jes´usAlcal´a-Fdez, Francisco Herrera, ―A New Multi objective 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] Salima Ouadfel , Souham Meshoul ," Handling Fuzzy Image Clustering with a Modified ABC Algorithm", *Intelligent Systems and Applications*,12,pp 65-74,2012.
[17] Dervis Karaboga , Bahriye Akay, A comparative study of Artificial Bee Colony algorithm, Elsevier, Applied Mathematics and Computation 214,pp 108–132,2009.