Association Rules Optimization using Artificial Bee Colony Algorithm with Mutation

Pankaj Sharma
Research Scholar, Computer Science Dept.
Vikrant Institute of Technology & Management
Gwalior, India

Sandeep Tiwari
Asst. Professor, Computer Science Dept.
Vikran Institute of Technology & Management
Gwalior, India

Manish Gupta
Asst. Professor, Computer Science Dept.
Vikrant Institute of Technology & Management
Gwalior, India

ABSTRACT
In data mining, Association rule mining is one of the popular and simple method to find the frequent item sets from a large dataset. While generating frequent item sets from a large dataset using association rule mining, computer takes too much time. This can be improved by using artificial bee colony algorithm (ABC). The Artificial bee colony algorithm is an optimization algorithm based on the foraging behavior of artificial honey bees. In this paper, artificial bee colony algorithm with mutation operator is used to generate high quality association rules for finding frequent item sets from large data sets. The mutation operator is used after the scout bee phase in this work. In general the rule generated by association rule mining technique do not consider the negative occurrences of attributes in them, but by using artificial bee colony algorithm (ABC) over these rules the system can predict the rules which contains negative attributes.

Keywords
Artificial bee colony (ABC), Mutation, Association rule, Support, Confidence, Frequent item set, Data mining.

1. INTRODUCTION
Now a days, growth of databases increase day-by-day due to the number of requirements or number of customers or end users. In this scenario, data mining [2, 9] plays an important rule for mining data according to the customer's requirements. Association rule mining is one of the popular and well known research methods for discovering interesting relationships between variables in large information repository or databases. Paper [1] describes how data mining and knowledge discovery are related to different fields like machine learning; statistics etc. genetic algorithm [5] based association rule finding is also discussed in paper [3]. For finding association rules, minimum support value plays an important rule. Paper [4] present a genetic algorithm based strategy for discovering association rules without specifying the value of minimum support. Many techniques have been proposed to optimize association rules [6, 7]. Also genetic algorithm based techniques [10, 11] have proposed previously.

In this paper, Artificial bee colony algorithm with mutation operator based association rule optimization technique has proposed.

The remainder of this paper is organized as follows. Section 2 describes the artificial bee colony algorithm with mutation. In section 3, Association rule mining is explained. Proposed algorithm is explained in section 4. Experimental results and parameter setup for result comparison are shown in section 5. Finally, section 6 concludes the paper.

2. ARTIFICAL BEE COLONY ALGORITHM WITH MUTATION
In paper [9, 11], one more phase in the form of mutation operator of genetic algorithm is added to original Artificial Bee Colony algorithm. In standard ABC algorithm, there are only 4 phases that described the overall working of this algorithm, but here one additional phase after the onlooker bee phase of ABC algorithm is added in the form of Mutation operator. Now modified artificial bee colony algorithm has five phases: first initialization phase, employed bee phase, onlooker bee phase, scout bee phase and finally mutation phase. For the local search, employed bee phase is used and the mutation phase is used to find out the new search area of the solution space. With the help of mutation mutation operator, there may be a possibility to change the local best position and the algorithm may not be trapped into local optima. In this work, the mutation phase is implemented on the probabilistic way in each iteration for searching food source during the life process of ABC optimization technique. Food sources are selected randomly from the food size to perform mutation operation. In mutation phase, if generated offspring's fitness is greater than the older one then replaces the older offspring's from the new one. Uniform mutation is used in this work.

The overall algorithm is described in following steps:

- Initialization phase.
- REPEAT

(a) In the Memory, Employed bees are placed on the food sources;

(b) Generate new offspring from older offspring after Applying mutation operator.

(c) In the memory, onlooker bees are placed on the food sources;

(d) For finding new food sources, Send the scout bee to the search space.

- UNTIL (requirements are not met).

3. ASSOCIATION RULE MINING
The main aim of association rule mining is to extract frequent item sets, correlation and association among different set of items in the transactional database, relational databases or other information repository.

Association rule mining algorithm finds association rules in the form of:

IF AB and CD then HELLO

IF UV and XY then BYE

Here AB, CD, UV and XY are different objects out of which if any person takes AB and CD then due to high propbility, he will take HELLO. Similarly if he will choose UV and XY then he will choose BYE.

In general, expressions which are in the form of A=>B, called association rules where A represents antecedent and B represents consequent.

Association rules represent how many times B has occurred if A has already occurred depending on the chosen support and confidence value. Here support is nothing but the probability of items or item sets in the given database (like transactional or other) and confidence represents conditional probability.

Apriori Algorithm:

In general, Apriori algorithm [8] works on two phases – first phase is to choose minimum support value which is applied in the database to find frequent item sets while in second phase, these item sets and the minimum confidence constraints are used to generate rules.

The pseudo code for the Apriori algorithm are given as follows -

Step 1: let C_n be the candidate item set of size n.

Step 2: let F_n be the frequent item set of size n.

Step 3: F_1 = {Frequent items}

Step 4: REPEAT

Step 5: C_{n+1} = Candidates generated from F_k ;

Step 6: REPEAT for each transaction t in database

Step 7: increment the count of all candidates in C_{n+1} that are contained in t.

Step 8 : F_{k+1} = Candidates in C_{n+1} with minimum support.

Step 9: UNTIL (F_n not equal to ϕ)

Step 10: return $U_n F_n$

4. PROPOSED METHODOLOGY

This section represents proposed methodology. Here ABC with mutation algorithm is applied over the rules gathered from apriori algorithm, to find frequent item sets.

In order to use the ABC algorithm with mutation, the following points must be addressed: initial population, fitness value, employed, onlooker, mutation and scout bees. Here Initial population is generated using randomly generated transactions. To calculate the fitness value of an individual, following fitness value is used-

$$f_i = \begin{cases} 1 / (1 + f_i) & \text{if } f_i >= 0 \\ 1 + \text{abs}(f_i) & \text{otherwise} \end{cases}$$

Other points are same like standard ABC algorithm, discussed above.

The steps of proposed algorithm for generating optimal association rules via ABC with mutation are as follows-

Step 1: Start

Step 2: Load dataset

Step 3: Find frequent item sets using apriori algorithm. Suppose F is the set of all frequent item sets generated by apriori algorithm and X is the output set, containing all generated association rules, initialized to zero.

Step 4: Set the termination condition for the ABC with mutation algorithm.

Step 5: Depict each item sets of Z and apply ABC with mutation algorithm on selected members to generate association rules.

Step 6: Evaluate fitness value of each rule.

Step 7: If the fitness function satisfied the desired criteria then add these rules in output set.

Step 8: if the desired number of generations not completed then goto step 3

Step 9: Stop

<u>*Block diagram of proposed work*</u>:

Block diagram of the proposed algorithm for optimizing association rules are given below and shown in figure 1.

5. EXPERIMENTAL RESULTS & PARAMETER SETUP

a. Data Sets :

To check the performance of the proposed work, different datasets are selected from UCI machine learning repository. Currently, 187 datasets are maintained by UCI machine learning research group. Out of these datasets, three popular datasets of Voting, Iris and Wine are selected for our experiments.

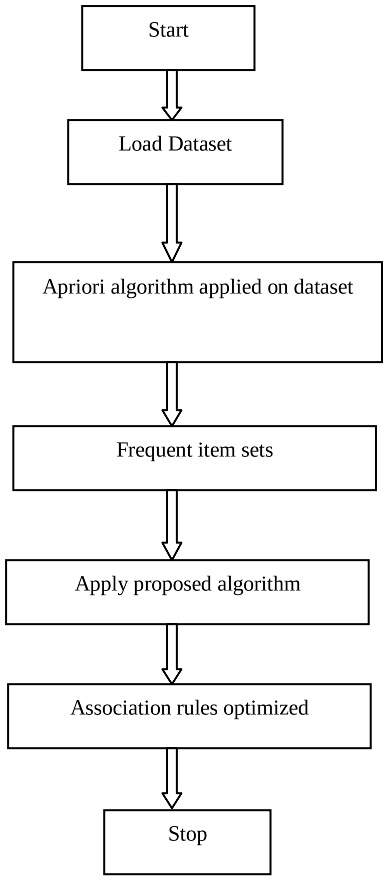

Fig 1 Shows the block diagram of proposed algorithm

Details of these datasets are given below –

- Voting dataset – Features =16
  Instances = 435
  Class = 2
- Wine dataset – Features =13
  Instances = 178
  Class = 3
- Iris dataset – Features =04
  Instances = 150
  Class = 3

b. Parameter Settings:

There are four main control parameter which are used to test the performance of proposed algorithm. First control parameter is the number of food sources which is equal to 20 and also it is equal to the number of employed bees and onlooker bees. Second control parameter is the maximum cycle number (MCN) which is equal to 2000 in this experiment. Third and the fourth control parameter are the mutation probability which is equal to 0.1 and the limit value. After the final rule has generated, two control parameter for class prediction are quality weight (α) and coverage weight (β), both are initialized to 0.5 in this experiment.

Proposed work is compared with KNN algorithm and standard ABC algorithm. Table 1 shows the performance classification accuracy. Figure 2 shows the graphical comparison between different algorithms.

Table 1: performance accuracy of classification

Datasets	***KNN (%)***	***ABC (%)***	***Proposed Work (%)***
Voting	95.10	97.21	98.19
Iris	94.08	96.44	97.44
Wine	96.22	98.13	98.39

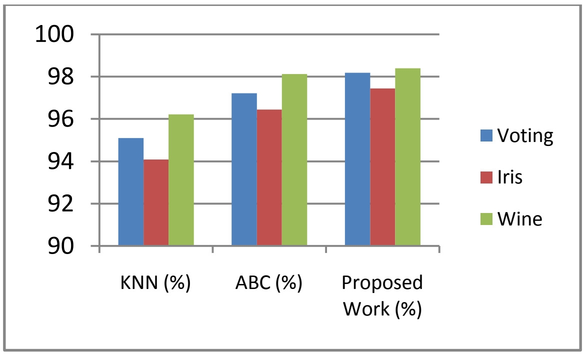

Figure 2 shows the performance of proposed work

6. CONCLUSION

Now days, the size of the databases are increased day–by-day. To find frequent item sets, there is a need of association rule mining. In this work, generated association rules using apriori algorithm are optimized using artificial bee colony with mutation algorithm, where mutation operator is used after the scout bee phase of ABC algorithm. To check the performance of proposed work, three datasets of voting, wine and iris are used, collected from UCI machine learning repository. Experimental results show that the performance of the proposed work with previously proposed works. Future work is to use the proposed work with different databases.

7. REFERENCES

[1] U. Fayyad and R. Uthurusamy, “Data Mining and Knowledge Discovery in Databases”, Communications of the ACM, vol. 39, no. 11, 1996, pp.24–34.

[2] J. Han and M. Kamber, “Data Mining Concepts and Techniques”, Morgan Kaufmann, 2006.

[3] W. Soto and A. Olaya-Benavides, "A Genetic Algorithm for Discovery of Association Rules." In Computer Science Society (SCCC), 2011, pp. 289-293.

[4] X. Yan, C. Zhang and S. Zhang, “Genetic Algorithm-Based Strategy for Identifying Association Rules without Specifying Actual Minimum Support”, Expert Systems with Applications, vol. 36, 2009, pp. 3066–3076.

[5] S. N. Sivanandamand and S. N. Deepa, “Introduction to Genetic Algorithms”, Springer-Verlag Berlin Heidelberg, 2008.

[6] M. Anandhavalli and S. Kumar Sudhanshu, A. Kumar and M.K. Ghose, “Optimized Association Rule Mining Using Genetic Algorithm”, Advances in Information Mining, vol. 1, issue 2, 2009, pp. 01-04.

[7] P. Wakabi-Waiswa and V. Baryamureeba, "Mining High Quality Association Rules using Genetic Algorithms", In Proceedings of the twenty second Midwest Artificial Intelligence and Cognitive Science Conference, 2009, pp. 73-78.

[8] Markus Hegland, “The Apriori Algorithm – a Tutorial”, CMA, Australian National University, WSPC/Lecture Notes Series, 22-27. March 30, 2005.

[9] Pujari A.K., “Data Mining Techniques”, Universities Press, 2001.

[10] S. Ghosh, S. Biswas, D. Sarkar and P.P. Sarkar, “ Mining Frequent item sets using Genetic Algorithm”, international journal of artificial intelligence & applications (IJAIA), Vol. 1, No. 4, October 2010.

[11] Mohit K. Gupta and Geeta Sikka, “ Association Rules Extraction using Multi-objective Feature of Genetic Algorihtm”, Proceedings of the world congresson engineering & computer science 2013 Vol II, WCECS 2013, 23-25 October- 2013, San Francisco, USA.

[12] Amit Singh, Neetesh Gupta and Amit Sinhal, “Artificial Bee Colony Algorithm with Uniform Mutation”, Proceeding of the international conference of soft computing for problem solving (SocPros 2011), Vol. 130, pp 503-511.