Improved Association Rules Optimization using Modified ABC Algorithm

Vineet Singh Bhadoriya
Research Scholar, Computer Science Dept.
Maharana Pratap College of Technology
Gwalior, India

Unmukh Dutta
Associate Professor, Computer Science Dept.
Maharana Pratap College of Technology
Gwalior, India

ABSTRACT
To discover the frequent item sets from the huge data sets, one of the most popular techniques of data mining, called association rule mining technique used. For generating association rules from huge database using association rule mining technique, Computer system takes too much. This can be enhanced, if the number of association rules generated using association rule mining technique from a huge dataset can be optimized. So here in this work, firstly association rules are generated using standard Apriori algorithm and then optimized these association rules using modified artificial bee colony (ABC) algorithm. In this modified ABC algorithm, one additional operator, called crossover operator, is used after the third phase, called scout bee phase, of ABC algorithm. Due to the better exploration property of crossover operator, it is used in this work. Experimental results show that the proposed schemes performance better than previously proposed schemes like K-Nearest Neighbor algorithm (KNN) and ABC algorithm.

Keywords
Artificial bee colony algorithm, ABC, Crossover operator, Association rules, Support, Confidence, Frequent item sets

1. INTRODUCTION
In today's scenario, the growth of databases increases regularly due to the online business, banking sector like businesses and other business where number of customers increase day-by-day. In this scenario, the mining of dataset from the large database according to customers as well as business requirements [2, 9] plays major role. Association rule mining technique is one of the most popular and well known data mining methods for finding interesting relationships between variables in large information repository or databases. How knowledge discovery and data mining are connected to different fields like statistics, machine learning etc in [1]. Genetic algorithm [5] based association rule mining from huge dataset is discussed in work [3]. For discovering association rules, the value of support plays important role. in work [4] present a new scheme based on a genetic algorithm for finding different association rules without specifying minimum support value. in literature, many techniques, either based on genetic algorithm [10, 11] or based on other optimization algorithm have been presented to enhanced association rules from a huge dataset [6, 7]. Paper [13] describes the genetic algorithm operator, called crossover operator, which is used in the particle swarm optimization algorithm. One new version of this work is proposed in [14] and modified version of ABC algorithm with mutation operator is discussed in [12].

In this work, modified artificial bee colony algorithm based association rule optimization technique has proposed, where both Apriori algorithm and ABC algorithm is used.

The remainder of this work is summarized as follows. Modified Artificial Bee Colony algorithm is discussed in section II. Section 3 described Association rule mining technique, where Apriori algorithm, is explained. Proposed work is described in section 4. Experimental results and parameter setup for checking the efficiency of proposed algorithm is explained in section 5. Finally, section 6 concludes the work.

2. MODIFIED ARTIFICAL BEE COLONY ALGORITHM
In this modified algorithm, one additional phase in the form of crossover operator of genetic algorithm is included in the standard Artificial Bee Colony algorithm. In artificial bee colony algorithm, only four phases described the overall working of the algorithm, but here one phase, called crossover operator phase, is added after the scout bee phase of the algorithm. Now ABC with crossover algorithm works in five different phases: first initialization phase then employed bee phase followed by the onlooker bee phase, the scout bee phase and finally crossover phase which is added in this work.

The steps of this modified algorithm are given below:

- Initialization phase.
- REPEAT

(a) Employed bees are placed on the food sources in the memory;

(b) Produce new chromosome from older chromosome after applying crossover operator.

(c) Onlooker bees are placed on the food sources in the memory;

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

(e) If crossover criteria met then generate new chromosome by replacing older one

- UNTIL (requirements not met).

3. ASSOCIATION RULE MINING
The major goal of the association rule mining is to generate frequent item sets, association and correlation among different set of relational database items, transactional database items or other information repository.

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

IF PQ and RS then HI
IF AB and CD then HELLO

Here PQ, RS, AB and CD are objects out of which if anyone takes PQ and RS then due to high probability, he will take HI. Similarly if he will select AB and CD then will choose HELLO.
In general, expressions which are in the form of P=>Q, called association rules where P represents antecedent and Q represents consequent.
Association rule represents how many times Q has occurred if P has already occurred in the database depending on the chosen confidence and support value. Here support value is nothing but the probability of item sets or different items in the given database (like relational, transactional or other database) and confidence value represents the conditional probability.

Apriori Algorithm:

In general, the Apriori algorithm [8] works on two different phases – first phase is to select the optimized support value which is used to discover frequent item sets from a huge database while in second phase, these frequent item sets and the minimum confidence value are used to produce association rules.

The pseudo code for the standard Apriori algorithm are described as follows -

Step 1: let C_n represents the different candidate item set of size n.
Step 2: let F_n represents the different frequent item set of size n.
Step 3: F_1 = {Frequent item sets}
Step 4: REPEAT
Step 5: C_{n+1} = Candidate item sets generated from F_k ;
Step 6: REPEAT for each transaction item 't' in the database
Step 7: increase the count of all candidate item sets in C_{n+1} that are contained in t.
Step 8: F_{k+1} = Candidate item sets having minimum support value.
Step 9: UNTIL (F_n not equal to ϕ)
Step 10: return $U_n F_n$

4. PROPOSED METHODOLOGY

This section described the proposed methodology. Here modified artificial bee colony algorithm is implemented over the association rules generated from the Apriori algorithm, to discover the frequent item sets from a huge dataset.

In order to use the modified artificial bee colony algorithm, the following terminologies must be addressed: initial population, fitness function and value, employed bee, crossover operator, onlooker bee and the scout bee. Here Initial population is formed using transactions which are taken randomly. To compute the fitness value, the following fitness function is used-

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

The basic steps of proposed work for producing optimal association rules via modified ABC algorithm are as listed below-

Step 1: Start

Step 2: Load standard dataset which are taken from internet

Step 3: Find different frequent item sets from huge database using the Apriori algorithm. Let F be the frequent item sets gathered from the Apriori algorithm and R containing all association rules generated from the Apriori algorithm, initialized to zero.

Step 4: Set the termination condition for the modified artificial bee colony algorithm.

Step 5: Show each item sets of R and apply modified ABC algorithm on selected items to generate optimized association rules.

Step 6: Estimate the fitness value of each association rule.

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

Step 8: if the desired different number of generations not accomplished then go to step 3

Step 9: Stop

Block diagram of proposed methodology:

Figure 1 shows the proposed methodology block diagram.

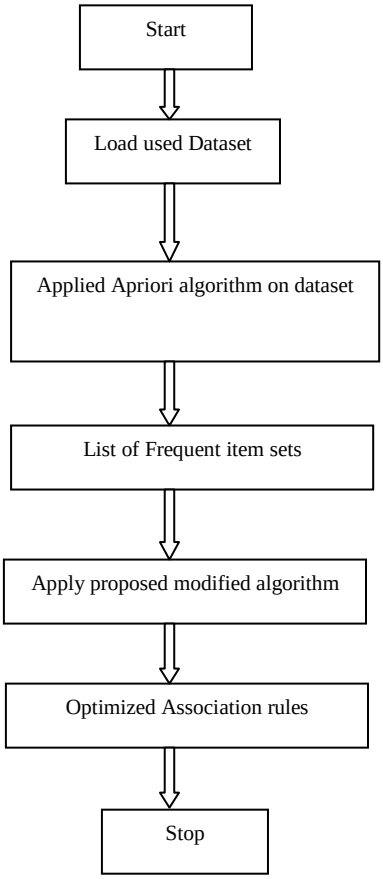

Fig 1: block diagram of proposed work

5. EXPERIMENTAL RESULTS & PARAMETER SETUP

5.1 Data Sets

To evaluate the efficiency of the proposed work, different datasets gathered from UCI machine learning repository. Currently, there are 187 datasets maintained by this UCI machine learning research group. Out of these different datasets, three most popular datasets of Iris, wine and Voting are selected for our experiments.

Details of these three datasets are given below

- Iris dataset –
  Number of Instances = 150
  Number of Classes = 3
  Total Features =04
- Wine dataset –
  Number of Instances = 178
  Number of Classes = 3
  Total Features =13
- Voting dataset –
  Number of Instances = 435
  Number of Classes = 2
  Total Features =16

5.2 Parameter Settings

To check the efficiency of proposed work with other algorithms, few control parameters are used which are listed below:

Control Parameter	Value
Number of Food Sources	20
Number of onlooker bees and employed bees	20
Maximum Cycle Number (MCN)	2000
Quality weight (α)	0.5
coverage weight (β)	0.5

Proposed work efficiency is compared with K-Nearest Neighbor algorithm and ABC algorithm. Table 1 shows the efficiency of classification while Figure 2 shows the pictorial comparison between proposed work and other different algorithms.

Table 1: Efficiency of Classification

Datasets	*KNN (%)*	*ABC (%)*	*Proposed Work (%)*
Voting	95.10	97.21	97.3
Iris	94.08	96.44	97.2
Wine	96.22	98.13	98.4

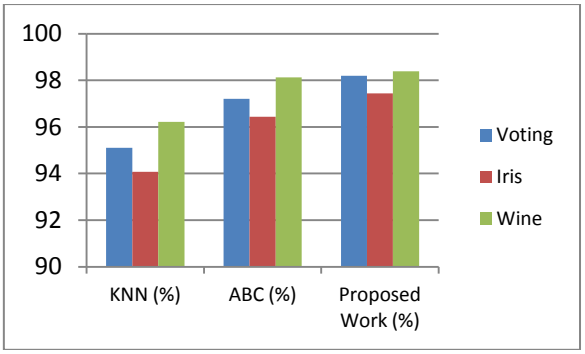

Figure 2: Efficiency of proposed work

6. CONCLUSION

Now days, the size and the number of databases are increases regularly due to the customers demand. To discover frequent item sets, association rule mining is used. In this work, firstly association rules are generated using the Apriori algorithm and then these rules are optimized using modified artificial bee colony algorithm. In this modified algorithm, one additional operator in the form of crossover is included after the scout bee phase of original ABC algorithm. To check the efficiency of proposed work, three datasets of iris, wine and voting are used, gathered from UCI machine learning repository. Experimental results show that the performance of the proposed work with previously proposed algorithms like KNN and ABC algorithm. Future work is to implement 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.

[13] Xie, Jiahua., Yang, Jie., "A Novel Crossover Operator for Particle Swarm Algorithm ", Machine Vision and Human-Machine Interface (MVHI), 2010 , IEEE Pages 161 – 164.

[14] Vineet S. Bhadoriya and Unmukh Dutta, "Association Rule Optimization using Artificial Bee Colony Algorithm with Crossover" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-5 Issue-1, March 2015.