Scientific Journal Impact Factor (SJIF): 5.278
International Journal of Modern Trends in Engineering and Research
IJMTER
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

Association Rule Mining Using Discrete Cuckoo Search Algorithm

Mali Trupti Rajendra
SSVPS, COE, DHULE

Abstract—Association Rule Mining is considered as one of the widely used data mining technique which allows users to find patterns in data. It has two steps firstly its mine frequent itemsets from dataset. Secondly it generate association rule from that frequent items. Many algorithms which is using for Association Rule Mining. Traditional Algorithms that mine too many association rules, which is not useful for users. Its degrading the computational efficiency. Recent research in Association Rule Mining it uses metaheuristic algorithms which is only looking for high quality of rule, a modified discrete cuckoo search algorithm for association rules mining DCS-ARM is proposed for that purpose. The effectiveness of algorithm is tested again the transactional dataset. The experimental result shown that the proposed algorithm outperforms BSO-ARM in terms of the number of valid rule and quality. DCS-ARM is less time consuming than BSC-ARM algorithm.

Keywords—Association Rule Mining; Basic Cuckoo Search; Discrete Cuckoo Search ;

I. INTRODUCTION

In a recent years, there is a lot of expeditious expansion in the field of rule mining due to the huge amount of data availability on the internet. Hence it becomes difficult to finding the specific pattern from the large data. To tackle this task a technique called association rule mining can be used. Association Rule Mining is important research area in data mining. The first task is mining association among set of item in transactional dataset. Association Rule Mining is a relation between two set of item X->Y, X is antecedent and Y is consequent part. Association Rule Mining it is a two stage process, in first stage its mining frequent items from dataset and in second stage it generate association rule from that frequent items. Association Rule quality measure by two measurements , is support and confidence. Support is an indication of how frequently the items appear in the data. Confidence indicates the number of times the if-then statements are found true. So many traditional algorithm which is useful for finding the association rules. It mines too many association rules, some rule which is not useful for the users, it mines invalid rule. For that reason, performance and complexity is a issue. For this reason, Metaheuristic algorithm are use for finding only high quality of rules. It uses a cuckoo search algorithm for finding high quality of rule. Is inspired by the obligate brood parasitic behaviour of some cuckoo species in combination with the Lévy flight behaviour of some birds and fruit flies. And another algorithm is discrete cuckoo search algorithm which outperform the basic cuckoo search algorithm, that mines high quality of rules than the Basic Cuckoo Search algorithm. Basic Cuckoo Search algorithm is limited in real number space. The main difference in between Basic cuckoo and discrete cuckoo algorithms is, First the nest is represent in binary variable and secondly Levy Flight is transforming into the change of probability .In Basic Cuckoo Search algorithm each nest position adjust according to the Levy Flights, and in Discrete Cuckoo Search algorithm the position which is adjusted by decomposition strategy. Both algorithms apply on dataset and finding high quality of rules.

II. RELATED WORK

Many variations of association rules mining algorithms have been proposed to improve the efficiency of mining association rules, as they have to mine a larger set of data items. Invalid rule they mine because of that it degrading the computational efficiency. However, the performance and complexity are the main issue. There are many algorithms which is useful for mining association rules. Some of the algorithm have been developed for rule mining .Let us discuss some of the previous algorithms which were developed.

In 2018, Rasha A. Mohammed and Mehdi G. Duaimi [1] proposed system, "Association rules mining using cuckoo search algorithm". They proposed Association Rule Mining using Cuckoo search algorithm . They have developed this system to finding association rule from the dataset. The objective was that to select a high quality of rule and improve the efficiency .

In 2014, Youcef Djenouri and Habiba Drias [2] proposed a system," Bees swarm optimisation using multiple strategies for association rule mining". This system uses Bees Swarm algorithm for finding association rule. It handle huge amount of data in web context. It uses three different strategy for improving the search area.

In 2013, K.N.V.D. Sarath,Vadlamani Ravi [3] proposed a work, "Association rule mining using binary particle swarm optimization". The work demonstrates that it using binary particle swarm optimization association rule miner for generating the association rules. It generate association rule without specifying minimum Support and minimum Confidence.

In 2012, Dongwoo Won and Dennis McLeod proposed a work, "An efficient approach to categorising association rules". This work [4] uses ontology for knowledge discovering from transactional data. It utilize the search space. It use hierarchical association rule categorization for finding needed rules efficiently.

In 2009, McLeod Xiaowei Yan a, Chengqi Zhang a, Shichao Zhang proposed a work, "Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support". This work [5] generate association rule by using genetic algorithm based strategy. It perform a global search and implemented a system automation because it never used user defined threshold value.

In 2004, Jiawei Han, Jian Pei, and Yiwen Yin proposed a work, "Mining Frequent Patterns without Candidate Generation". This work [6] uses Fp-growth algorithm for mining association rule. It uses divide and concurred strategy .It has two scan in fist scan it produce frequent items. Then FP-tree generated.In second scan it mines the FP-tree for each item.

In 1993, Rakesh Agrawal Tomasz Imielinski Arun Swami proposed a work, "Mining Association Rules between Sets of Items in Large Databases". This work [8] uses Apriori algorithm for mining association rule. It has a two stages, in first stage it generate frequent itemsets, and in second stage it generate association rule from that frequent items.

In 1993, Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami proposed a work, "Data Mining: A Performance perspective". This work [7] uses AIS algorithm for generate all itemset in transactional database. An estimation is used in the algorithm to prune those candidate itemsets that have no hope to be large. It is suitable for low cardinality sparse transaction database.

III. METHODOLOGY

Association Rule mining is one of the research area in data mining. It plays an various important role such as clustering, classification, prediction, and association analysis. Finding all frequent items is more time consuming because finding various pattern from massive amount of data. It is very time consuming for user it only focus on subset of useful rule. Many association rules mining algorithms have been proposed for improve the efficiency of mining association rules, as they have to mine a larger set of data items. However, the performance and complexity of association rules mining algorithms are main issue. Many metaheuristic algorithms which is useful for solving such problem. The main reason for using metaheuristic algorithm is that it is not like other algorithms, It only focus for finding the best rule. For this purpose, a new algorithm based on discrete CS metaheuristic is using for finding high-level association rules.

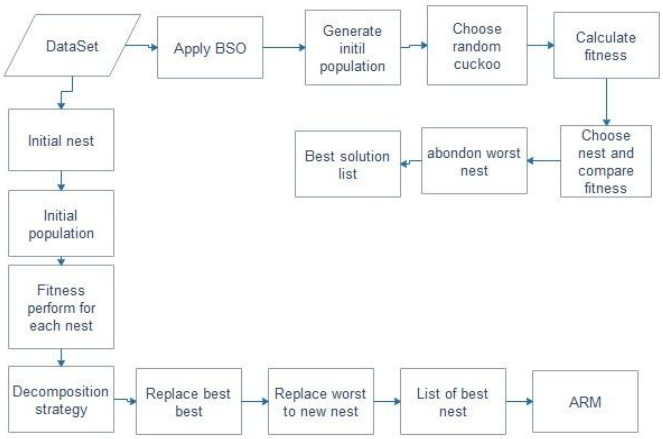

Figure 1 Architecture of the Rule Mining

It is started with choosing the algorithm which is apply on dataset. Firstly apply BSO-ARM algorithm it produce initial population, from that population choosing random nest via levy flight. Then calculate its fitness and again choosing the random nest and compare fitness of that two nest. Best nest is choose. Abondon the worst Pa nest. Produced the best solution list. Secondly choosing DCS-ARM algorithm which is apply on dataset. It produce initial nest randomly generate initial population from that initial nest. For each nest in population calculate fitness. Perform decomposition strategy and getting two nest position Compare the fitness And replace nest with best fitness nest. Worst nest are remove and replace it with new solution. Save the best nest in the List of Best Nests; Generate the association rule

3.1. Basic Cuckoo Search Algorithm

Cuckoo Search algorithm is metaheuristic algorithm which was proposed by Xin-She Yang and Saush Deb in 2009 For solving the various obtimisation problem.This algorithm which is inspired by the behavior of some cuckoo species the behavior is called Levy Flight behavior.[9]

Algorithm begins with an primary population, a group of cuckoos. These cuckoos lay some eggs in the habitat of other host birds. A random group of potential solutions is generated that represent the habitat in Cuckoo Algorithm. In the fitness function will be evaluated the parameters of the candidate components. The steps to search the optimum solution are given as follows. First, the algorithm begins with a primary population of birds and they have some eggs to lie in some host bird's habitats. Some of these eggs grow up and become adult birds which are more like to the host bird's eggs and other eggs are discovered by host birds and are removed. The more profit is gained in that area, the more eggs that remain and hatch in the place. So the place in which more eggs remain

will be the term that Cuckoo Algorithm is going to optimize. Each cuckoo has a "distance" to the best habitat.

```
Begin
Objective function f(x), x = (x1, ..., xd)T
Generate initial population of n host nests xi (i = 1, 2, ..., n)
While (t < MaxGeneration) or (stop criterion)
Get a cuckoo randomly by Lévy flights
Evaluate its quality/fitness Fi
Choose a nest among n (say, j) randomly
If (Fi > Fj ),
Replace j by the new solution;
End
Fraction (pa) of worse nests is abandoned, and new ones are built;
Keep the best solutions (or nests with quality solutions);
Rank the solutions and find the current best
End while
Postprocess results and visualisation
End
```

3.2. Discrete Cuckoo Search Algorithm

Basic Cuckoo Search Algorithm is limited in real number space. For such reason Discrete Cuckoo Search Algorithm is proposed. Two main differences between Basic cuckoo search algorithm and discrete cuckoo search algorithm is, nest are represented in a binary variable and second is the Levy Flight which is transform into the change of probability. [10]

```
Input: β, pa, MaxGeneration, a, b, Transactional Dataset, Minsup, Minconf
Output: Set of Asso. Rules
Xi ←, the initial nest, generated randomly
While (t < MaxGeneration) or (stop criterion)
Begin
Generate the initial population of n host nests from Xi
For each X in population
Evaluate its fitness Fx
Get Xa and Xb from X by decomposition strategy
Evaluate fitness of Xa (Fxa)and Xb (Fxb)
If Fx is greater than Fxa and Fxb
X is not replaced
Else If Fxb greater than Fxa
Replace X by Xb
Else
Replace X by Xa
End if
End for
Fraction (pa) of worse nests are deserted, and new ones are built // where pa refers to
Save the best nest in the List of Best Nests;
Xi← the best nest in List of Best Nests // best of the best
t ← t + 1
End while
For each nest X in List of Best Nests do
```


25 Generate the rule from S
26 End for

IV. PERFORMANCE EVALUATION AND GRAPH

A dataset of zoo is the input to the system. The output of the system is trend to show how there is to generate the high quality of Association Rule. The input dataset has been downloaded from the link. https://dtai.cs.kuleuven.be/CP4IM/datasets In this system firstly select the algorithm which is apply on dataset.

Fistly select the Basic Cuckoo Search algorithm. During the Basic cuckoo search algorithm, the algorithm perform via levy flight, levy flight which is basically use for new solution generation. After initialization, it reading the train data, then it creating host nest randomly via leavy flight,and calculate fitness then it sorted to the best fitness ,generate initial population of n host

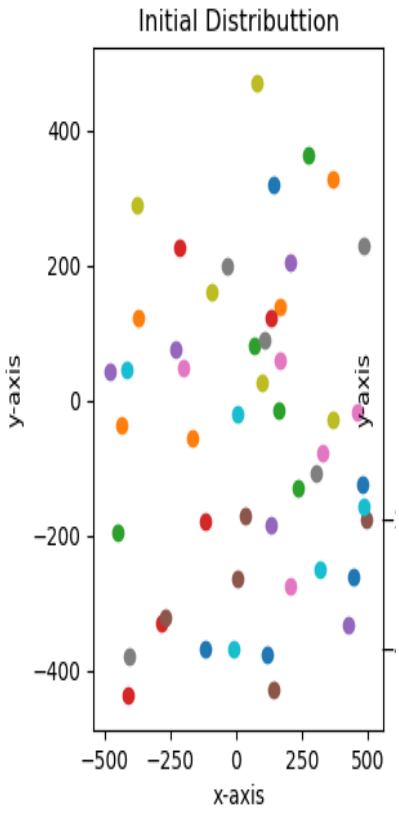

Figure 2 Initial Population

After initial distribution, it finds the new solution via levy flight from that solution choosing a random cuckoo, finding its fitness and then compare the fitness values of nest and replace old solution with new one this loop is continuously perform until termination criteria is satisfied it perform 1000 iteration and generate its fitness value for choosing the best solutions.

After iterations from that fraction pa of worse nest in abandoned, and new one are build. Keep the best solution from that solutions. Rank the solutions and find the current best return the optimize nest and generate the graph for best solutions

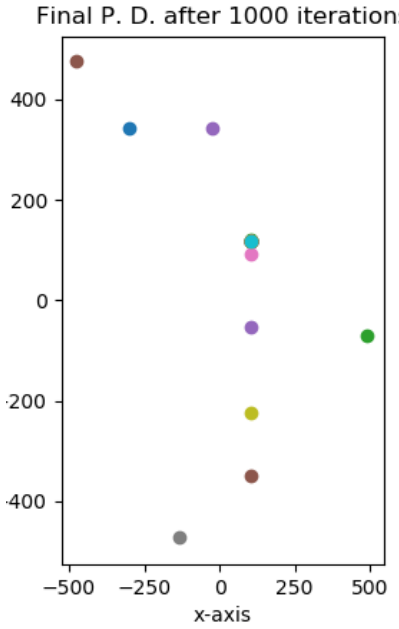

Figure 3 Final Distribution

During the DCS-ARM algorithm, which is apply on dataset.It genearate initial nest, from initial nest it generate initial population. For each nest finding the fitness getting two possible position of nest by decomposition strategy. Evaluate fitness of two possible position of nest. Compare the fitness values, then it replace better fitness position value. Fraction pa of worst nest are deserted and new nest are build. Save the best nest in the list of best nest. and Generate association rule from that best nest list.

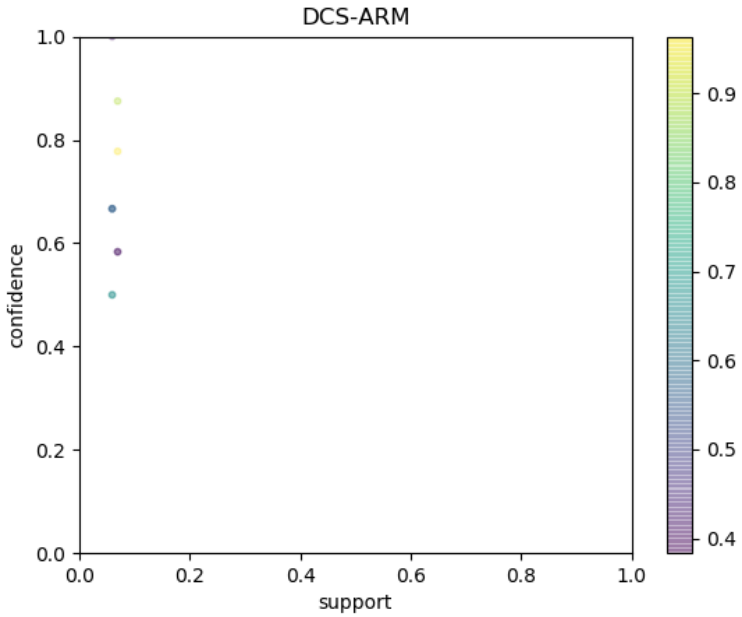

Figure 4 DCS-ARM Graph

V. CONCLUSION

The system reduce the unnecessary rule. The DCS-ARM outperforms BSO-ARM in terms of the number of valid rule and quality. DCS-ARM is less time consuming than BSC-ARM algorithm.

REFERENCES

[1] R. A. M. a. M. G. Duaimi, "Association rules mining using cuckoo search," *Association rules mining using cuckoo search*, vol. 10, no. 1, p. 16, 2018.

[2] Y. D. H. a. a. Z. Djenouri, "Bees swarm optimisation using multiple strategies for," *Int. J. Bio-Inspired Computation*, vol. 6, no. 4, p. 239–249., 5 2014.

[3] K. a. R. V. Sarath, "'Association rule mining using binary particle swarm," *Engineering Applications of Artificial Intelligence,*, vol. Vol. 26, no. 8, p. pp.1832–1840., 2013.

[4] D. a. M. Won, "An efficient approach to categorising association rules," *Int. J. Data Mining, Modelling and Management,*, vol. 4, no. 4, p. 309–333., 2 2012.

[5] C. Z. a. S. Z. b. Xiaowei Yan a, "Genetic algorithm-based strategy for identifying association," *Proceeding of 2009 World*, vol. 36, no. 1, p. 11, 2009.

[6] J. P. a. Y. Y. Han, *Data Knowledge and Knowledge Discovery*, vol. 8, no. 1, pp. 53-87, 2 2004.

[7] T. I. a. A. N. S. A Rakesh Agrawal, "Data Mining: A performance prespective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 5, no. 6, pp. 914-925., 1993.

[8] "Mining Association Rules between Sets of Items in Large Databases," in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data,*, San Jose, 1993

[9] X. a. D. S. Yang, "Cuckoo search via Levy flights," in Proceeding of 2009 World, 2009.

[10] H. e. a. Zheng, "A discrete cuckoo search algorithm for solving knapsack problems," International Journal on Advances in Information Sciences and Service Sciences, vol. 4, no. 1, pp. 331-339, 3 2012.

