Optimization of Association Rule Mining Apriori Algorithm Using ACO

Badri Patel, Vijay K Chaudhari, Rajneesh K Karan, YK Rana

***Abstract*— Association rule mining is an important topic in data mining field. In a given large database of customer transactions. Each transaction consists of items purchased by a customer in a visit. Apriori algorithm that generates all significant association rules between items in the database. On the basis of the association rule mining and Apriori algorithm, this paper proposes an improved algorithm based on the Ant Colony Optimization algorithm. We can optimize the result generated by Apriori algorithm using Ant colony optimization algorithm. The algorithm improved result produces by Apriori algorithm. Ant Colony Optimization (ACO) is a metaheuristic inspired by the foraging behavior of ant colonies. ACO was introduced by Dorigo and has evolved significantly in the last few years.**

***Index Terms*— Association rule mining, Apriori algorithm, Ant Colony Optimization (ACO) algorithm, data mining**

I. INTRODUCTION

Data mining refers to extracting knowledge from large amounts of data. Data mining is often treated as synonym for another popularly used term, Knowledge Discovery in Databases (KDD)[2].Several organizations have collected massive amounts data. These data sets are usually stored on storage database systems. One of the main reasons for the limited success of database systems in this area is that current database systems do not provide necessary functionality for a user interested in taking advantage of this information. This paper introduces the problem of "mining" a large

collection of basket data type transactions for association rules between sets of items with Some minimum specified confidence, and presents an efficient algorithm. For this purpose, using the Rough Sets approach, one can deal with two major problems in the analysis of an information system:

- Reducing unnecessary objects and attributes so as to get the minimum subset of attributes, ensuring a good approximation of classes and an acceptable quality of classification.
- Representing the information system as a decision table which shows dependencies between the minimum subset of attributes(called conditions) and particular class numbers(called decisions), without redundancy. The organization of the rest of the paper is as follows.

In Section 2, we present Introduction of Apriori algorithm and ACO, In Section 3, we discuss related work. In Section 4, we present our algorithm for optimization of association rules mining. We conclude with a summary in Section 6.

II. INTRODUCTION OF APRIORI ALGORITHM

Let D the task-relevant data, be a set of database transactions where each transaction T is a set of items, called Tid. Let I= {I1, I2,…, Im} be a set of items. An item set contains k items is a k- item set. If a k-item set satisfies minimum support (Min_sup) then it is a frequent k-item set, denoted by Lk. Firstly Apriori algorithm generated a set of candidates, which is candidate k-item sets, denoted by Ck. If the candidate item sets satisfies minimum support then it is frequent item sets. The description of the algorithm is given below:

1. Suppose a minimum support threshold Min_sup) and a minimum confidence threshold (Min_conf) [8]
2. Scan the dataset, candidate 1-itemsets, C1, and the number of occurrences of each item is determined. The set of frequent 1-itemsets, L1, is then determined, consisting of those candidate 1-itemsets in C1 having minimum support. The algorithm uses L1 ∞ L1 to generate candidate 2-itemsets, C2.
3. Scan the dataset again, frequent 2-itemsets, L2, is then determined, consisting of those candidate 2-itemsets in C2 having minimum support. Candidate 3-itemsets, C3 is then generated by L2 ∞ L2.
4. Repeatedly scan the dataset, compare the support count of each candidate in Ck-1 with Min_sup, and then generate Lk-1 , join Lk-1 ∞ Lk-1 to generate Ck until no more candidate item sets.

A two-step process is used to find the frequent itemsets: join and prune actions.

a) The join step: To find Lk, Ck is generated by joining Lk-1 with itself if member l1 and member l2 are joined.

Manuscript received March 2011.

Badri Patel, M.Tech., Department of IT, Technocrat Institute of Technology, Bhopal(M.P.), Email: bpatel.it@gmail.com

Prof. Vijay K Chaudhari, Head, Department of Information Technology, Truba Institute of Technology, Bhopal (M.P.), Email: vijay_ashish@yahoo.com

Prof. Rajneesh K Karan, Dean Academic, Radharaman Engineering College, Bhopal (M.P.)

Prof. Y.K. Rana, Head, Department of Computer Science and Engineering & Informationa Technology, Radharaman Engineering College, Bhopal (M.P.).

b) The prune step: The members of Ck may not be frequent. A scan of the database to determine the count of each candidate in Ck, and use Lk-1 to remove a candidate k-item set from Ck would result in the determination of Lk.

In many cases the Apriori candidate generate-and-test method reduces the size of candidate sets. However if mining a large set of database, the Apriori algorithm will produce overfull candidates o frequent itemsets, so the algorithm needs scan database frequently when finding frequent itemsets. And it will take more resource and time to accomplish one scanning. So it must be inefficient.

III. INTRODUCTION OF ACO ALGORITHM

ACO algorithms were inspired from natural behavior of ant colonies. ACO has been applied successfully to numerous hard optimization problems including the traveling salesman problem (TSP) [2]. Artificial ants are simple agents implementing constructive heuristics. The basic idea of constructive

heuristics is incrementally construct solutions by adding, in each step, a solution component to a partial solution until to a complete solution is formed. The cooperation is the key element of ACO algorithms once good solutions are resulted of the cooperative interaction of several artificial ants during the construction of solutions.

ACO has been applied to a broad range of hard combinatorial problems. Problems are defined in terms of components and states, which are sequences of components. Ant Colony Optimization incrementally generates solutions paths in the space of such components, adding new components to a state. The ACO system contains two rules:

1. Local pheromone update rule, which applied whilst constructing solutions.
2. Global pheromone updating rule, which applied after all ants construct a solution.

Furthermore, an ACO algorithm includes two more mechanisms: trail evaporation and, optionally, daemon actions. Trail evaporation decreases all trail values over time, in order to avoid unlimited accumulation of trails over some component. Daemon actions can be used to implement centralized actions which cannot be performed by single ants, such as the invocation of a local optimization procedure, or the update of global information to be used to decide whether to bias the search process from a non-local perspective [6][7].

At each step, each ant computes a set of feasible expansions to its current state, and moves to one of these in probability. The probability distribution is specified as follows. For ant k, the probability of moving from state t to state n depends on the combination of two values [8]:

- The attractiveness of the move, as computed by some heuristic indicating the priori desirability of that move;
- The trail level of the move, indicating how proficient it has been in the past to make that particular move: it represents therefore an a posteriori indication of the desirability of that move.

IV. RELATED WORK

1. Yan-hua WANG Xia FENG 'The Optimization of Apriori Algorithm Based on Directed Network', (2009 Third International Symposium on Intelligent Information Technology Application) this paper gives an experiment to analyze and compare the difference between the two (Apriori algorithm and proposes an improved algorithm based on the directed network) algorithms and the result shows that the improved algorithm promotes the efficiency of computing. In this paper, algorithm improved that based on directed network.
2. Yiwu Xie, Yutong Li, Chunli Wang, Mingyu Lu" The Optimization and Improvement of the Apriori Algorithm", Through the study of Apriori algorithm we discover two aspects that affect the efficiency of the algorithm. One is the frequent scanning database, the other is large scale of the candidate item sets. Therefore, Apriori algorithm is proposed that can reduce the times of scanning database, optimize the join procedure of frequent item sets generated in order to reduce the size of the candidate item sets. In this paper It not only decrease the times of scanning database but also optimize the process that generates candidate item sets.

V. PROPOSED WORK

This work presents an ACO algorithm for the specific problem of minimizing the number of association rules. Apriori algorithm uses transaction data set and uses a user interested support and confidence value then produces the association rule set. These association rule set are discrete and continues therefore weak rule set are required to prune. Optimization of result is needed.

To optimize these we are going to propose Ant colony optimization (ACO) algorithm for association rule optimization. Using the confidence value as the pheromone (p) value and compute the path updating value (P)=P+Δt

Where $\Delta t = (2^{d+1}-1)/d$, (d-Number of transaction set), then optimize association rule set is generated.

VI. CONCLUSION

We have proposed in this paper an ACO algorithm for optimization association rule generated using Apriori algorithm. This work describes a method for the problem of association rule mining. An ant colony optimization (ACO) algorithm is proposed in order to minimize number of association rules.

REFERENCES

[1] Yan-hua,Wang, Xia Feng,'The optimization of Apriori algorithm based on directed network. 3rd international Symposium on intelligent information technology application, 2009.

[2] J Han, M Kamber,' Data mining: Concepts and

techniques', Morgan Kaufman Publishes, 1992.

[3] Gao, Shao-jun Li,' A method of improvement and optimization on association rules appriori algorithm', proceeding of the 6th cogress on intelligent control and automation,2006 pp5901-5905

[4] Estefan G.M de L Manoel atl.,' Minimum number of switching operations via ant colony optimization', 19th internation conference on electricity distribution Vienna, 21-24 May2007.

[5] R. Agrawal,T. Imielinski, A. Swami,' Mining Association rules between sets of items in large database,' proceeding of 1993 ACM SIGMOD conference Washington DC, USA.

[6] AK Jain, MN Murthy, PJ Flynn,' Data clustering- A review', ACM Computing Surveys, Oct. 2001

[7] M Dorigo , T Stutzle, 'Ant colony optimization', The MIT press Cambridge, MA.

[8] Karla Taboada, S Mabu, E Gonzales,' Genetic Network programming for fuzzy association rule based classification, 2009.

[9] Karla Taboada,Shingo Mabu, Eloy Gonzales," Genetic Network Programming for Fuzzy Association Rule-Based Classification", 2009