ISSN 2348 - 2249
Volume-3, Issue-1, January-March, 2015
International Journal for Multi Disciplinary Engineering and Business Management
(IJMDEBM)

ARM Using Ant Colony Optimization Algorithm - A Review

Ritu Walia
Department of Computer Science
Mata Gujri College
Fatehgarh Sahib, India
de.ritu@yahoo.co.in

Abstract **-** Association rule mining represents a promising technique to find hidden patterns in large data bases. Ant Colony Optimization (ACO) is a metaheuristic approach for solving hard combinatorial optimization problems. ACO is inspired by shortest path searching behavior of various ant species. This paper presents the optimization efficiencies of the ant colony algorithms used for data mining according to requirement of the application.

Keywords **-** Data Mining, Association Rules, Apriori Algorithm, ACO

I. INTRODUCTION

Data mining is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. Data mining is about solving problems by analyzing data already present in databases [1]. Data is the raw fact. Processed data is called information. Fact of knowing about the world is called knowledge. Example- Cotton Produces Cloth. Cotton is the raw fact is called data. Produces by using some machine i.e. data in processing state is called information. The final output cloth is the knowledge. Knowledge is closely related with Intelligence. A person having more knowledge is called highly intelligence person [2].

The term Knowledge Discovery in Databases (KDD) is generally used to refer to the overall process of discovering useful knowledge from data, where data mining is a particular step in this process. The additional steps in the KDD process, such as data cleaning, integration, selection, transformation, Data Mining, pattern evaluation and knowledge representation of the results of the data mining process, ensure that useful knowledge is derived from the data [3].

The Knowledge Discovery in Databases process comprises of a few steps leading from raw data collections to some form of new knowledge. The iterative process consists of the following steps:

Data cleaning: it is a phase in which noise of data and irrelevant data are removed from the collection.

Data integration: at this stage, multiple data sources, often heterogeneous, may be combined in a common source.

Data selection: at this step, retrieved only the relevant data according to the user need.

Data transformation: also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.

Data mining: it this step, clever techniques are applied to extract patterns potentially useful.

Pattern evaluation: in this step, strictly interesting patterns representing knowledge are identified based on given measures.

Knowledge representation: is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results.

II. ASSOCIATION RULES

Agrawal and Srikant first proposed the issue of the association rule mining in 1993. Association rule mining is used to find the relationship between the items. An association rule is in the form of X→Y, where X and Y represent Itemsets (I) [5].

Each association rule has two measures of rule interestingness, Support and Confidence.

Support of a rule A⇒ B is the probability of the itemset {A, B}. This gives an idea of how often the rule is relevant.

$$\text{Support}(A \Rightarrow B) = P(\{A, B\}), \quad (1)$$

Confidence of a rule A⇒ B is the conditional probability of B given A. This gives a measure of how accurate the rule is.

$$\text{Confidence}(A \Rightarrow B) = P(B|A) = \text{support}(\{A, B\}) / \text{support}(A), \quad (2)$$

For example:

Computer ⇒ financial_ management_ software [Support=2%, Confidence=60%] (3)

A support of 2% for association Rule (3) means that 2% of all the transactions show that computer and financial management software are purchased together. A confidence of 60% means that 60% of the customers who purchased a computer also bought the software.

The aim of association rule is to find all association problems having support and confidence not less than given threshold value. For the given support i.e. min_sup, if the item set of D's support is not less than min_sup, then it can say that D is the frequent item set [6].

ISSN 2348 - 2249
Volume-3, Issue-1, January-March, 2015
International Journal for Multi Disciplinary Engineering and Business Management
(IJMDEBM)

III. Apriori Algorithm

Apriori Algorithm was developed by R. Agarwal and R. Srikant in 1994. Apriori employs an iterative approach known as a level-wise search, where k-itemsets are used to explore (k+1)-itemsets, to mine frequent itemsets from transactional database for Boolean association rules [7].

First, the set of frequent 1-itemsets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support. The resulting set is denoted L1. Next, L1 is used to find L2, the set of frequent 2-itemsets, which is used to find L3, and so on, until no more frequent k-itemsets can be found. The finding of each Lk requires one full scan of the database. To improve the efficiency of the level-wise generation of frequent itemsets, an important property called the Apriori property, presented is used to reduce the search space [7].

Apriori property: All nonempty subsets of a frequent itemset must also be frequent.

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

Join Step: Ck is generated by joining Lk-1with itself.

Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset.

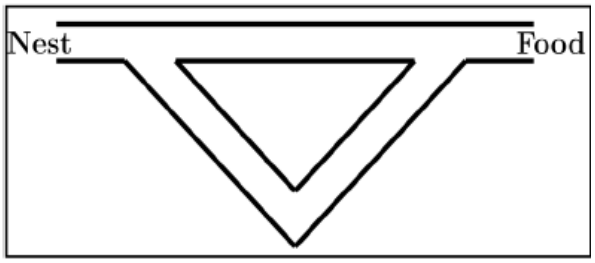
IV. Ant Colony Optimization

Ant Colony Optimization (ACO) is a metaheuristic for solving hard combinatorial optimization problems. It was proposed by Marco Dorigo in 1991. The inspiring source of ACO is the pheromone trail laying and following behavior of real ants. These ants deposit pheromone on the ground in order to mark some favorable path that should be followed by other members of the colony. Ant colony optimization exploits a similar mechanism for solving optimization problems [16]. The first example of such an algorithm is Ant System (AS), which was proposed using as example application the well-known traveling salesman problem (TSP) [8].

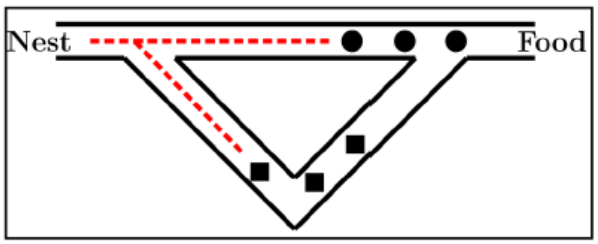
How ants can find shortest paths between food sources and their nest?

When searching for food, ants initially explore the area surrounding their nest in a random manner. While moving, ants leave a chemical pheromone trail on the ground. Ants can smell pheromone. When choosing their way, they tend to choose, in probability, paths marked by strong pheromone concentrations. As soon as an ant finds a food source, it evaluates the quantity and the quality of the food and carries some of it back to the nest. During the return trip, the quantity of pheromone that an ant leaves on the ground may depend on the quantity and quality of the food. The pheromone trails will guide other ants to the food source [9].

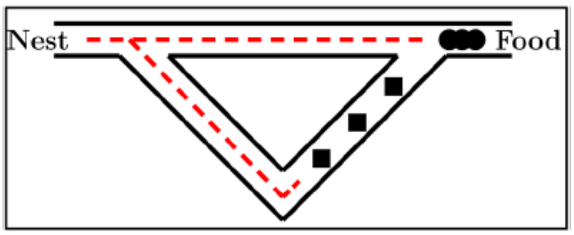
Example:



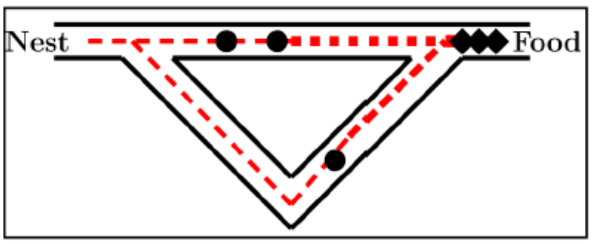
(a) All ants in the nest. There is no pheromone in the enviroment.



(b) The foraging starts. In probability, 50% of ants take the short path (symbolized by circels) and 50% of ants take the long path to the food source (symbolized by rhombs).



(c) The ants that have taken the short path have arrived earlier at the food source. Therefore, when returning, the probabilty to take again the short path is higher



(d) The pheromone trail on the short path recieves, in probabilty, a stronger reinforcement, and the probabilty to take this path grows. Finally, due to the evaporation of the pheromone on the long path, the whole colony will, in probabilty, use the short path.

Fig 4.1 An experimental setting that demonstrates the shortest path finding capability of ant colonies. Between the ants' nest and the only food source exist two paths of different lengths. In the four graphics, the pheromone trails are shown as dashed lines whose thickness indicates the trails' strength.

The ACO system contains two rules:

1. Local pheromone update rule, which applied whilst constructing solutions.

ISSN 2348 - 2249
Volume-3, Issue-1, January-March, 2015
International Journal for Multi Disciplinary Engineering and Business Management
(IJMDEBM)

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 [10].

V. Related Work

Hybrid algorithm was proposed to solve combinatorial optimization problem by using Ant Colony and Genetic programming algorithms [10]. Evolutionary process of Ant Colony Optimization algorithm adapts genetic operations to enhance ant movement towards solution state. Ants deposit a certain amount of pheromone on the components; that is, either on the vertices or on the edges that they traverse. The amount of pheromone deposited may depend on the quality of the solution found. Ants adaptively modify the way the problem is represented and perceived by other ants, but they are not adaptive themselves. The genetic programming paradigm permits the evolution of computer programs which can perform alternative computations conditioned on the outcome of intermediate calculations, which can perform computations on variables of many different types, with the ability to perform iterations and recursions to achieve the desired result, and which can define and subsequently use computed values and sub-programs, and whose size, shape, and complexity is not specified in advance.

Paper [11] presents a promising technique to find hidden patterns in large data bases. The main issue in a medical data is that large number of rules can be discovered, most of which are irrelevant, which makes the speed of searching slow and, also not all of the generated rules are interesting, and some need to be ignored. In medical terms, association rules related to disease data measures the patient risk factors and occurrence of the disease. Association rule medical significance is evaluated with the usual support and confidence metrics. Authors proposed an algorithm that mines the association among the various attributes in a dataset. They conducted experiments on synthetic and real data sets. For future work authors suggest to conduct experiments on large real time health datasets to predict the diseases like heart attack and compare the performance of algorithm with other related algorithms.

Association rule mining is one of the most popular data mining methods. When mining the association rules, large number of rules then select only the interesting rules are found. In this paper, authors present the comparison of Apriori and FP-growth algorithms. The performance is analyzed based on the execution time for different number of instances and confidence in Super market data set. The efficiency of both algorithms is evaluated based on time to generate the association rules. Authors concluded that the FP-growth algorithm behaves better than the Apriori algorithm [12].

In this paper, Authors present a swarm intelligence based technique for mining rules over a medical database. Rules are a suitable method for representing real world medical knowledge because of their simplicity, uniformity, transparency, and ease of inference. Swarm Intelligence (SI) has been applied to the rule mining process as its dynamic nature provides flexibility and robustness to process of rule mining [13]. The rule quality can be viewed in terms of its accuracy and comprehensibility. A rule will be interesting to a medical practitioner if it is accurate and easily understood. A system generating large number of rules or rules with too many conditions in the antecedent tends to confuse the end user and is not usable for medical knowledge discovery. ACO/PSO with new quality measure of fitness performs the best in terms of comprehensibility and accuracy. This method also penalizes false positives severely, which is a desirable property for data mining in the medical domain. One drawback of the approach is the complexity of the algorithm.

The Probabilistic Traveling Salesman Problem (PTSP) is a TSP problem where each customer has a given probability of requiring a visit. The goal is to find an a priori tour of minimal expected length over all customers, with the strategy of visiting a random subset of customers in the same order as they appear in the a priori tour. Authors were investigated the potentialities of ACO algorithms for the PTSP. In particular, they have shown that the pACS algorithm is a promising heuristic for homogeneous TSP instances. Moreover, for customer's probabilities close to 1, the ACS heuristic is a better alternative than pACS. At present they were investigating the heterogeneous PTSP, for different probability configurations of customers. This is an interesting direction of research, since it is closer to a real-world problem than the homogeneous PTSP. Authors were also trying to improve pACS performance by inserting in the ants' tour construction criterion information about the customers probabilities. Work which will follow this paper also comprehends a comparison of pACS with respect to other PTSP algorithms [14].

The management of reactive resources plays an important role in maintaining voltage stability and system reliability. This paper [15] presents a new method to find the optimal solution to reactive regulation in power system, using the daily data collected in power substations. The new algorithm is combined with improved ant colony algorithm and Apriori data mining technique. In this paper, Apriori algorithm has been improved and applied to substation data mining process. Ant colony algorithm is applied to get the optimal solution of reactive power allocation in substations. The state transition probability formula is amended and parameters are dynamically adjusted in this ant colony algorithm. The choice of the ant's path to the next node is determined by the tabu table formulated according to the confidence level of the data mining. The switching strategy of the capacitor sets are given by online algorithm. An example substation system is described to test the algorithm proposed in this paper. Experimental results show that, reactive power optimization

ISSN 2348 - 2249
Volume-3, Issue-1, January-March, 2015
International Journal for Multi Disciplinary Engineering and Business Management
(IJMDEBM)

method based on data mining system can improve the system efficiency, reduce power loss, and have a great significance of stable operation.

Number of rules generated by using authors approach is comparatively less than the number of rules generated by the Apriori algorithm and less than MOGA (Multi Objective Genetic Algorithm). The main area of concentration in this paper [17] is to optimize the rules generated by Association Rule Mining (Apriori Algorithm), using hybrid evolutionary algorithm. The main idea for using Evolutionary algorithms in the discovery of high-level prediction rules is that they perform a global search and cope better with attribute interaction than the greedy rule induction algorithms often used in data mining. The future enhancements will be on using the other Evolutionary Optimization Algorithms such as PSO (Particle Swarm Optimization) for the rule generation.

After review of different techniques on Association Rule Mining Using Ant Colony Optimization Algorithm. We comparison the various techniques of Ant Colony Optimization Algorithm in the table below:

Table 5.1 Comparison the various techniques of Ant Colony Optimization Algorithm.

Authors	**Techniques**	**Purpose**	**Conclusion/Future Scope**
Nada M. A. Al Salami [10]	Genetic Algorithm, ACO	Hybrid algorithm is proposed to solve combinatorial optimization problem by using Ant Colony and Genetic programming algorithms.	Evolutionary process of ACO algorithm adapts genetic operations to enhance ant movement towards solution state. The algorithm converges to the optimal final solution, by accumulating the most effective sub-solutions.
K.Srinivas et al. [11]	Decision trees, minimum support, minimum confidence	Authors proposed an algorithm that mines the association among the various attributes in a dataset.	For future work authors suggest to conduct experiments on large real time health datasets to predict the diseases like heart attack and compare the performance of algorithm with other related algorithms.
K.Vanitha, R.Santhi [12]	Apriori Algorithm, FP-Tree Algorithm	Authors present the comparison of Apriori and FP-growth algorithms.	Authors concluded that the FP-growth algorithm behaves better than the Apriori algorithm.
Veenu Mangat [13]	ACO, Particle swarm optimization, rule quality.	In this paper, present a swarm intelligence based technique for mining rules over a medical database.	One drawback of the approach is the complexity of the algorithm.
Leonora Bianchi et al. [14]	ACO	The goal is to find an a priori tour of minimal expected length over all customers, with the strategy of visiting a random subset of customers in the same order as they appear in the a priori tour.	pACS should be improved by adding to the tour construction phase a local search algorithm. The best choice and design of such a local search is also an interesting issue for the PTSP.
Gong Jinxia, et al. [15]	Apriori Algorithm, ACO	This paper [15] presents a new method to find the optimal solution to reactive regulation in power system, using the daily data collected in power substations.	Experimental results show that, reactive power optimization method based on data mining system can improve the system efficiency, reduce power loss, and have a great significance of stable operation.
J.Arunadevi, Dr.V.Rajamani [17]	Evolutionary Optimization, Algorithms, Genetic Algorithms, ACO	To reduce the association rules.	In future also use the other Evolutionary Optimization Algorithms such as PSO (Particle Swarm Optimization) for the rule generation.

Gives the uses and limitations of various algorithms in the below table:

Table 5.2 Uses and limitations of various algorithms.

Algorithms	**Uses**	**Limitations**
Apriori Algorithm	a) It uses large itemset property. b) It is easily parallelized. c) It is easy to implement.	a) It Assumes transaction database is memory resident. b) It requires many database scans.

FP-Tree Algorithm	a) In this only 2 passes over data-set. b) It compresses data-set. c) In this no candidate generation. d) It is much faster than Apriori.	a) FP-Tree may not fit in memory. b) FP-Tree is expensive to build.
Ant Colony Optimization (ACO)	a) It can be used in dynamic applications. b) Positive Feedback leads to rapid discovery of good solutions. c) Distributed computation avoids premature convergence. d) For TSPs (Traveling Salesman Problem), relatively efficient.	a) Convergence is guaranteed, but time to convergence uncertain b) Coding is not straightforward
Genetic Algorithm (GA)	a) It can quickly scan a vast solution set. b) Bad proposals do not affect the end solution negatively as they are simply discarded. c) The inductive nature of the GA means that it doesn't have to know any rules of the problem-it works by its own internal rules. d) This is very useful for complex or loosely defined problems.	a) Genetic Algorithm involves longer running times on the computer. Fortunately, this disadvantage continues to be minimized by the ever-increasing processing speeds of today's computers.

VI. DATA SOURCE

For the current study the data deployed will be accumulated from both the fieldwork as well as the paperwork. The fundamental data that will hold the power to influence the conclusion would be the information fetched from the respondents via well framed questionnaire. Further, a pilot study in the region of North India will also be pursued so as to ensure the flexibility and extensibility of the responses collected via questionnaire. The questionnaire will be drafted by keeping in focus the objectives of the research study and the structure of questionnaire will be in direct congruence with the present research literature and the discussions held with the banking personnel, bank customers and the respective academicians. However, wherever required the secondary data may also be deployed to complete the present research study.

VII. 8. CONCLUSION

This paper demonstrates the review of ant colony algorithm based data mining techniques. ACO algorithm is used for optimized the association rules. Good quality rules help in better decision making. The ACO Algorithm can be used for generating good quality rules mined from Apriori Algorithm. The modified Association Rule Mining Process will help in generating even better quality of rules.

REFERENCE

[1] Lan H.Witten,Eibe Frank,Mark A. Hall, “Data Mining Practical Machine Learning Tools and Techniques”, Third Edition, Morgan Kaufmann Publishers is an imprint of Elsevier 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA.

[2] Arabinda Nanda1 Saroj Kumar Rout2, “Data Mining & Knowledge Discovery in Databases: An AI Perspective”, Proceedings of national Seminar on Future Trends in Data Mining (NSFTDM-2010):-10th may, 2010, Organenised by Department of Computer Science, Gandhi Engineering college, Bhubaneswar.

[3] Joyce Jackson “Data Mining: A Conceptual Overviewdata Mining: A Conceptual Overview”, Communications of the Association for Information Systems (Vol. 8, 2002) 267-296.

[4] Osmar R. Zaïane,” Introduction to Data Mining”, CMPUT690 Principles of Knowledge Discovery in Databases”, University of Alberta, Department of Computing Science, 1999.

[5] Manisha Gupta, “Application of Weighted Particle Swarm Optimization in Association Rule Mining”, International Journal of Computer Science and Informatics (IJCSI) ISSN (PRINT): 2231 –5292, Vol. 1, Issue-3.

[6] Ms. Sanober Shaikh, Ms. Madhuri Rao, “Association Rule Mining Based On Trade List”.

[7] Jiao Yabing,” Research of an Improved Apriori Algorithm in Data Mining Association Rules”, International Journal of Computer and Communication Engineering, Vol. 2, No. 1, January 2013.

[8] Marco Dorigo and Thomas Stiutzle, “Ant Colony Optimization: Overview and Recent Advances”, IRIDIA, Universit´e Libre de Bruxelles (ULB), Brussels, Belgium, M. Gendreau, J. Y. Potvin (eds.), Handbook of Metaheuristics, International Series in Operations Research & Management Science 146, DOI 10.1007/978-1-4419-1665-5 8, Springer Science Business Media, LLC 2010.

[9] Christian Blum, "Ant colony optimization: Introduction and recent trends", Physics of Life Reviews 2 (2005) 353–373, Available online at www.science direct.com.
[10] Nada M. A. Al Salami, "Ant Colony Optimization Algorithm", UbiCC Journal, Vol. 4, No. 3, August 2009.
[11] K.Srinivas, G.Raghavendra Rao and A.Govardhan, "Mining Association Rules from Large Datasets Towards Diseas Prediction", 2012 International Conference on Information and Computer Networks (ICICN 2012) IPCSIT vol. 27 (2012)
[12] K.Vanitha, R.Santhi, "Evaluating The Performance Of Association Rule Mining Algorithms", Journal of Global Research in Computer Science, Vol. 2, No. 6, June 2011.
[13] Veenu Mangat, "Swarm Intelligence Based Technique for Rule Mining in the Medical Domain", International Journal of Computer Applications (0975 – 8887), Vol. 4 – No.1, July 2010.
[14] Leonora Bianchi, Luca Maria Gambardella, and Marco Dorigo, "An Ant Colony Optimization Approach to the Probabilistic Traveling Salesman Problem", PPSN VII, LNCS 2439, pp. 883–892, Springer-Verlag Berlin Heidelberg 2002
[15] Gong Jinxia, Xie Da, Zhang Yanchi, Jiang Chuanwen, "Combination of Data Mining and Ant Colony Algorithm for Reactive Power Optimization", 2011 Third International Conference on Measuring Technology and Mechatronics Automation.
[16] Marco Dorigo, Mauro Birattari, and Thomas St¨utzle, "Ant Colony optimization", IEEE Computational Intelligence Magazine | November 2006 1556 603X/06/$20.00, 2006IEEE.
[17] J.Arunadevi, Dr.V.Rajamani, "Optimization of Spatial Association Rule Mining using Hybrid Evolutionary algorithm", International Journal of Computer Applications (0975 – 8887.

