An Improved Association Rule Mining Approach to Reduce Iterations in Ant Colony Algorithm through Artificial Bee Colony Algorithm

Girish Kumar Ameta
M.Tech Scholar,
Arya College of Engineering & Information Technology, Jaipur, Rajasthan, India

Dr. Vibhakar pathak
Professor
Department of Computer Science & Engineering,
Arya College of Engineering & Information Technology, Jaipur, Rajasthan, India

Abstract- In data mining, Associated Rules are working to find the best search data from the large amount of data. There are number of different algorithm available to find the associated rules like Ant colony algorithm, Apriori and FP growth and may more. In this Research, we find the problem for find best cost in low number of the iterations. As the number of iterations wills decreases, the performance of the system will increase and also reduces the steps for scanning the large database.

Keywords – Ant Colony Algorithm (ACO), Artificial Bee Colony (ABC), MCN, SN

I. INTRODUCTION

Information mining is assuming vital part in current developing rate of web information. It is additionally a crucial field of research in the field of example extraction and social event of data on given database. The assignment of information mining is to separate valuable learning for human clients from a database. There are number of various apparatuses accessible with the end goal of examining the information. It permits clients to examine information from very different adequacy, characterize it, and outline the connections recognized. In this manner in specialized way, it is the way toward discovering connections or examples among many fields in substantial social databases.[1]

Numerous computerized reasoning and measurable techniques exist for information investigation and elucidation. Nonetheless, these techniques were frequently not intended for the huge data sets information mining is managing today. Terabyte sizes are normal. This raises the issues of adaptability and productivity of the information mining strategies when handling impressively expansive information. Calculations with exponential and even medium-arrange polynomial multifaceted nature can't be of functional use for information mining. Straight calculations are generally the standard. In same topic, examining can be utilized for mining rather than the entire dataset. Nonetheless, concerns, for example, fulfillment and selection of tests may emerge. Different points in the issue of execution are incremental upgrading, and parallel programming. There is most likely parallelism can take care of the size issue if the dataset can be subdivided and the outcomes can be consolidated later. Incremental overhauling is critical for blending comes about because of parallel mining, or upgrading information mining comes about when new information gets to be distinctly accessible without having to re-examine the entire dataset

II. ASSOCIATION RULES

Association rule mining is the conclusive system in the field of data mining. It goes for extricating fascinating relationship, visit example, affiliation or easygoing structure among set of thing in the exchange database or other information vaults. Association mining is utilized as a part of different ranges for instance banking, retail establishments and so on. Association control mining has an extensive variety of pertinence; for example, advertise wicker bin examination, suspicious email location, library administration and numerous territories. The ordinary calculation of affiliation standards disclosure continues in two stages. All continuous thing sets are found in the

initial step. The regular thing set is the thing set that is incorporated into at any rate least bolster exchanges. The affiliation rules with the certainty in any event least sure are created in the second step. [4]
In general, the association rule is an expression of the form X=>Y, where X is antecedent and Y is consequent. Association rule shows how many times Y has occurred if X has already occurred depending on the support and confidence value. Many algorithms for generating association rules were presented over time. Some well-known algorithms are Apriori, FP-Growth, Ant Colony Optimization and Genetic algorithm.[5]

Every association rule has two quality estimations, Support(S) and Confidence(C) (Yan et al., 2009). Support of an administer A ⇒ B is the likelihood of the itemset {A, B}. This gives a thought of how regularly the govern is applicable: Support (A ⇒ B) = P ({A, B}), Confidence of a control A ⇒ B is the contingent likelihood of B given A. This gives a measure of how precise they manage it. Certainty (A ⇒ B) = P (B|A) = Support ({A, B})/Support (A).

A. Basic Concepts –

Let I = {I1, I2, … , Im} be an arrangement of m unmistakable qualities, T be exchange that contains an arrangement of things to such an extent that T ⊆ I, D be a database with various exchange records Ts. An association rule is a suggestion as X⇒Y, where X, Y ⊂ I are sets of things called itemsets, and X ∩ Y =∅. X is called predecessor while Y is called subsequent, the manage implies X infers Y. By and large, an arrangement of things, (for example, the precursor or the resulting of a lead) is called an itemset. The quantity of things in an itemset is known as the length of an itemset. Itemsets of some length k are alluded to as k-itemsets. By and large, an affiliation rules mining calculation contains the accompanying strides: 1. The arrangement of applicant k-itemsets is created by 1-augmentations of the vast (k - 1)- itemsets produced in the past cycle. 2. Bolsters for the competitor k-itemsets are produced by an ignore the database. 3. Itemsets that don't have the base support are disposed of and the rest of the itemsets are called expansive k-itemsets. This procedure is rehashed until not any more huge itemsets are found.

III. ANT COLONY OPTIMIZATION

Ant Colony Optimization (ACO) (Vittorio Maniezzo et al.) is a paradigm for designing metaheuristic algorithms for combinatorial optimization problems. This algorithm was first proposed by M. Dorigo, 1992 [9]. Ant Colony Algorithm is a multi-agent approach for solving difficult combinatorial optimization problems like Traveling Salesman, vehicle routing, sequential ordering, graph coloring, routing in communications networks [9][10].

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.

Moreover, an ACO calculation incorporates two more systems: trail dissipation and daemon activities. Trail vanishing diminishes all trail values after some time, so as to maintain a strategic distance from boundless collection of trails over some part. Daemon activities can be utilized to execute brought together activities which can't be performed by single ants, for example, the conjuring of a nearby improvement method, or the redesign of worldwide data to be utilized to choose whether to predisposition the hunt procedure from a non-neighborhood point of view. [11][12].

Ant Colony Optimization (ACO) is inspired by shortest path searching behavior of various ant species. Ants are social insects that live in colonies and because of their mutual interaction they are capable of performing difficult tasks. A very interesting aspect of behavior of ant species is their ability to find shortest path between the ants' nest and the food sources [11]. Ants (blind) go through the food while laying down pheromone. Shortest path is discovered via pheromone trails;

- Each ant moves at random.
- Pheromone is deposited on path on which ant moves.
- More the pheromone trails better the path (positive feedback sys).
- Ants follow the intense pheromone trails.

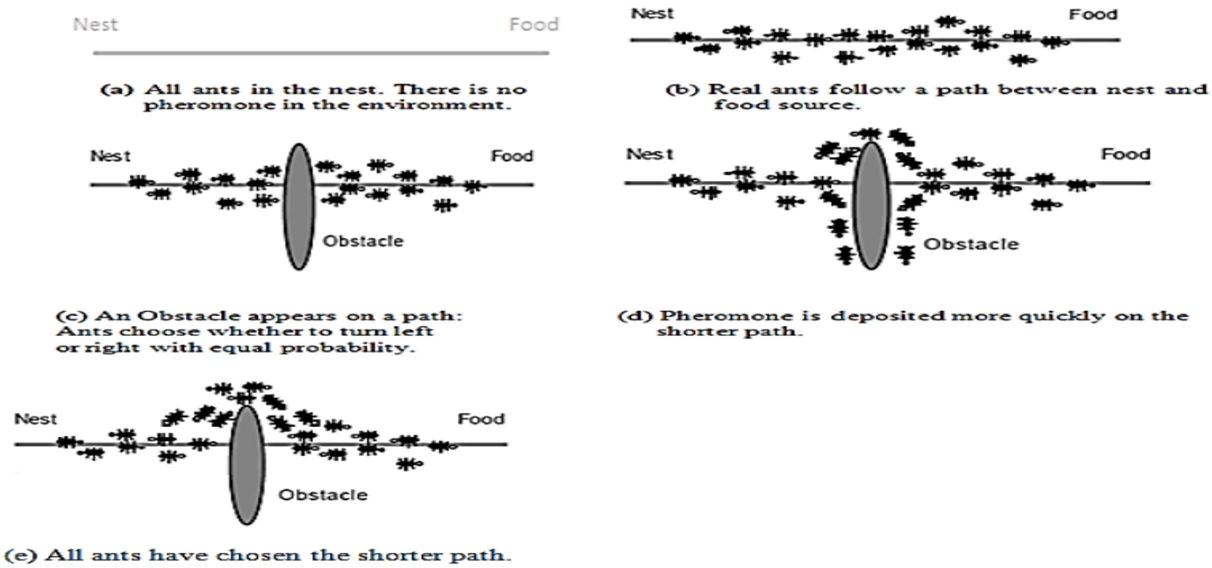

Figure 3.1: An experimental setting that demonstrates the shortest path finding capability of ant colonies. Between the ants' nest and the only food source exists two paths of different lengths.

IV. ARTIFICIAL BEE COLONY ALGORITHM

ABC calculation, the state of simulated honey bees comprises of three gatherings of honey bees: utilized honey bees, spectators and scouts. To start with half of the province comprises of the utilized manufactured honey bees and the second half incorporates the spectators. For each nourishment source, there is just a single utilized honey bee. At the end of the day, the quantity of utilized honey bees is equivalent to the quantity of nourishment sources around the hive. The utilized honey bee that's the sustenance source has been relinquished by the honey bees turns into a scout.

In ABC calculation, the position of a nourishment source speaks to a conceivable answer for the streamlining issue and the nectar measure of a sustenance source relates to the quality (wellness) of the related arrangement. The quantity of the utilized honey bees or the spectator honey bees is equivalent to the quantity of arrangements in the populace.[12]

At the initial step, the ABC creates a haphazardly disseminated introductory populace P(G = 0) of SN arrangements (nourishment source positions), where SN signifies the measure of populace. Every arrangement xi (i = 1, 2, ..., SN) is a D-dimensional vector. Here, D is the quantity of enhancement parameters. After instatement, the number of inhabitants in the positions (arrangements) is subjected to rehashed cycles, C = 1, 2, ...,MCN, of the inquiry procedures of the utilized honey bees, the spectator honey bees and scout honey bees. An utilized honey bee delivers a change on the position (arrangement) in her memory relying upon the nearby data (visual data) and tests the nectar sum (wellness esteem) of the new source (new arrangement). Given that the nectar measure of the new one is higher than that of the past one, the honey bee remembers the new position and overlooks the old one. Else she keeps the position of the past one in her memory[13]. After every single utilized honey bee finish the inquiry procedure, they share the nectar data of the nourishment sources and their position data with the passerby honey bees on the move zone. A passerby honey bee assesses the nectar data taken from every single utilized honey bee and picks a sustenance source with a likelihood identified with its nectar sum. As on account of the utilized honey bee, she creates an alteration on the position in her memory and checks the nectar measure of the applicant source. Providing that its nectar is higher than that of the previous one, the bee memorizes the new position and forgets the old one. An artificial onlooker bee chooses a food source depending on the probability value associated with that food source, *pi* , calculated by the following expression (1):

$$P_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \qquad \ldots\ldots (4.1)$$

where fit_i is the fitness value of the solution *i* which is proportional to the nectar amount of the food source in the position *i* and *SN* is the number of food sources which is equal to the number of employed bees (*BN*). In order to produce a candidate food position from the old one in memory, the ABC uses the following expression (2):

$$v_{i,j} = x_{i,j} + \emptyset_{i,j}(x_{i,j} - x_{k,j}) \quad \ldots..(4.2)$$

where $k \in \{1, 2,..., SN\}$ and $j \in \{1, 2,...,D\}$ are randomly chosen indexes. Although k is determined randomly, it has to be different from i. $\varphi_{i,j}$ is a random number between [-1, 1]. It controls the production of neighbor food sources around $x_{i,j}$ and represents the comparison of two food positions visually by a bee. As can be seen from (4.2), as the difference between the parameters of the $x_{i,j}$ and $x_{k,j}$ decreases, the perturbation on the position $x_{i,j}$ gets decrease, too. Thus, as the search approaches to the optimum solution in the search space, the step length is adaptively reduced. If a parameter value produced by this operation exceeds its predetermined limit, the parameter can be set to an acceptable value. In this work, the value of the parameter exceeding its limit is set to its limit value. The food source of which the nectar is abandoned by the bees is replaced with a new food source by the scouts. In ABC, this is simulated by producing a position randomly and replacing it with the abandoned one. In ABC, providing that a position cannot be improved further through a predetermined number of cycles, then that food source is assumed to be abandoned. The value of predetermined number of cycles is an important control parameter of the ABC algorithm, which is called "limit" for abandonment. Assume that the abandoned source is x_i and $j \in \{1, 2,...,D\}$, then the scout discovers a new food source to be replaced with $x_{i.}$[13][14]

V. PROPOSED METHOD

In this research, through below mentioned work flow, we are trying to find the best cost value in low iterations. Data mining is the process in which associated rules are extracted from the large amount of data.

1. Consider no of iterations as 200 including all initial parameters.
2. Consider an empty bee solution; initialize the population array and best solution.
3. Create the initial population with a null array to hold the best cost values.
4. Start the main loop of ABC.
5. Define acceleration coefficient with new bee position update.
6. Calculate the cost function for new bee.
7. Compare and update the population values.
8. Calculate the fitness values by using mean cost.
9. In the internal loop of onlooker bee compare and update the best cost value.
10. End of the main loop of ABC.
11. Compare both best cost values with respect to iterations.

VI. RESULTS

6.1 is showing the best coast value for the ACO algorithm (Ant Colony Optimization). As we can see from the Figure the value of the best coast is not getting constant up to 1000 Iterations. As the numbers of iterations get decrease the value of the best get is also get decrease. From the Ant Colony algorithm it is clear that the Best cost value is not constant after long iterations also.

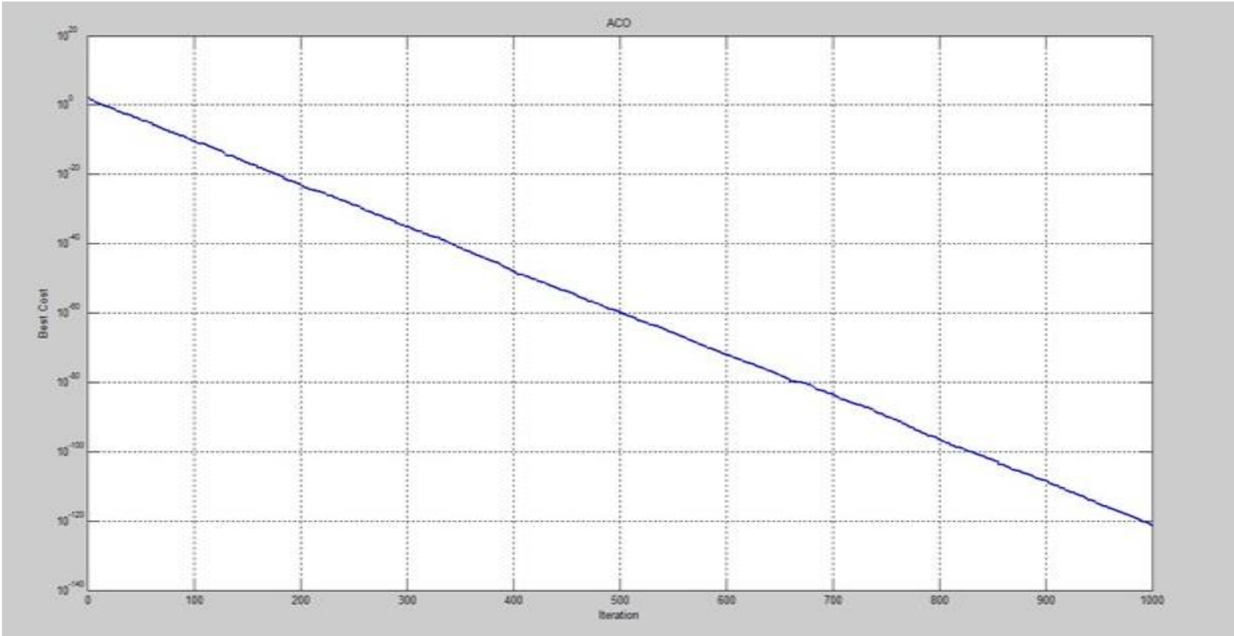

Figure 6.1:- Best Cost for ACO

Figure 5.2 is showing the best coast value for the proposed algorithm ABC (Artificial Bee Colony Algorithm) . As we can see from the figure that the best cost value is getting constant at the few iterations. Near about at 180 iterations it is getting constant which is very low from the ACO algorithm iterations.

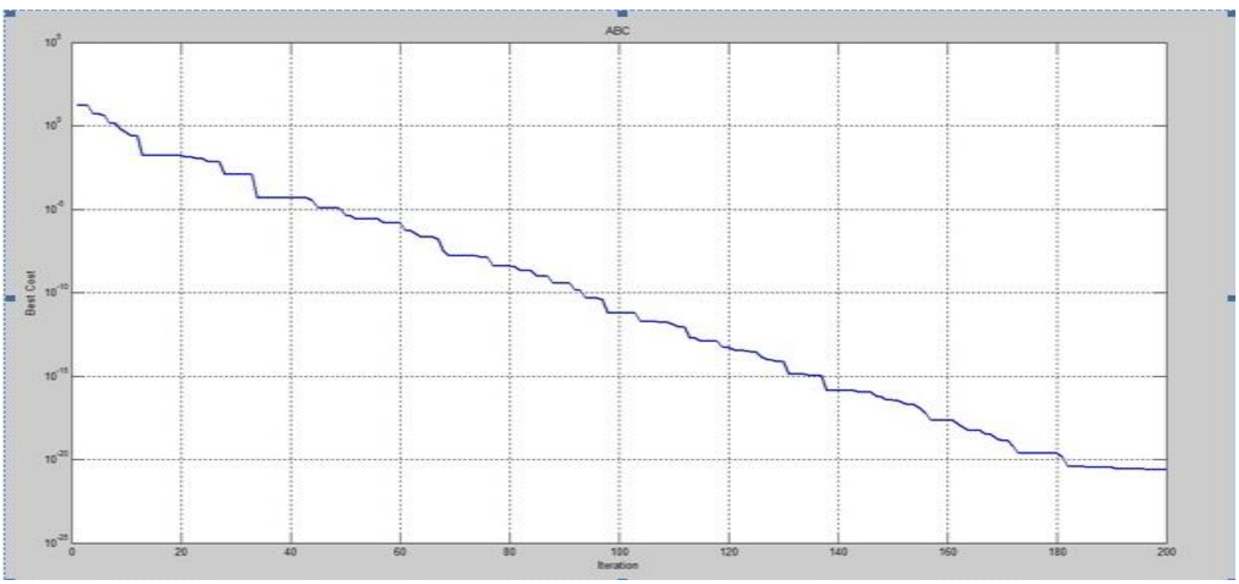

Figure 6.2 :- Best Cost for ABC (Artificial Bee Colony Algorithm)

Figure 6.3 is showing the output waveform comparison for both the Algorithms ABC and ACO . As we can see from the Figure 6,3 the comparison of the ACO and ABC algorithm . ABC algorithm is using low iterations for give the constant best value of the system as compare to the ACO Algorithm.

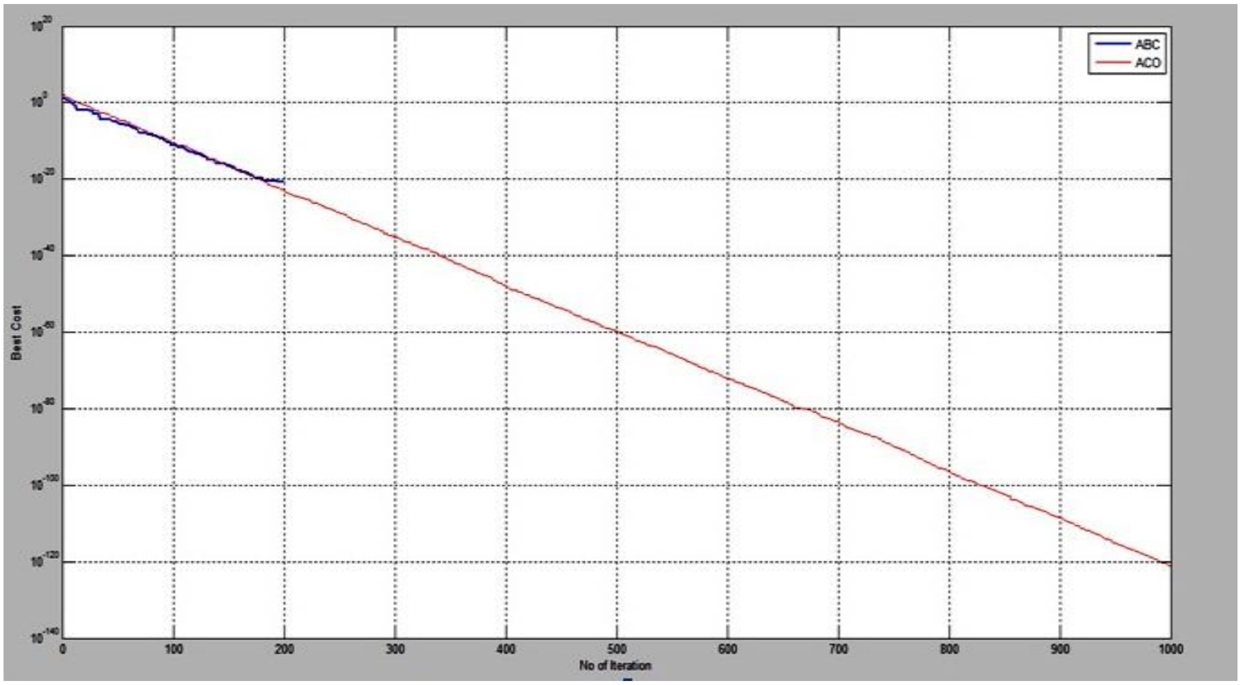

Figure 6.3:- Best Cost for ABC (Artificial Bee Colony Algorithm) and ACO (Artificial Bee Colony Algorithm)

VII. CONCLUSION

Association Rule Mining is the essential part of data mining process. Apriori Algorithm is the popular algorithm of association rule mining. Apriori Algorithm that generates all significant association rules between items in the database. On the basis of the association rule mining and Apriori algorithm, an improved algorithm based on the Artificial Bee Colony(ABC) algorithm is proposed. We can optimize the result generated by ACO(Ant Colony Optimization) algorithm using Artificial Bee Colony(ABC) Algorithm by introducing Probabilistic Scheme. The algorithm improves result produces by ACO algorithm. As we can see from the results session , ABC is using low number of iterations to get the best value as compare to ACO (Ant Colony Optimization) .In the last , we can say that for this research of Data Mining , ABC is superior as compare to ACO(Ant Colony optimization) for find the best value in the low iterations .

REFERENCES

[1] Arabinda Nanda and Saroj Kumar Rout, Data Mining & Knowledge Discovery in Databases: An AI Perspective, Proceedings of National Seminar on Future Trends in Data Mining (NSFTDM-2010), Organized by Department of Computer Science, Gandhi Engineering college, Bhubaneswar, 10th may, 2010.

[2] Jing He, Advances in Data Mining: History and Future, Third international Symposium on Information Technology Application, 978-0-7695- 3859-4, IEEE, 2009

[3] Meenakshi Sharma, Data Mining: A Literature Survey, International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-3, Issue-2).

[4] K.Saravana Kumar, R.Manicka Chezian, A Survey on Association Rule Mining using Apriori Algorithm ,in *International Journal of Computer Applications (0975 – 8887), Volume 45– No.5, PP- 47-50 , May 2012.*

[5] Yan et al (2009), Genetic Algorithm-Based Strategy For Identifying Association Rules Without Specifying Actual Minimum Support, Available online at www.sciencedirect.com, Expert Systems with Applications 36,3066–3076.

[6] Sotiris Kotsiantis, Dimitris Kanellopoulos (2006), Association Rules Mining: A Recent Overview, GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), pp. 71-82

[7] Theresa Barker (2005), Meggie von Haartman, Ant Colony Optimization, IEEE 516 Spring.

[8] Karaboga D, An idea based on honey bee swarm for numerical optimization, Erciyes University, Turkey, Technical report TR06

[9] M. Dorigo, V. Maniezzo, A. Colorni (1991) ,The ant system: an autocatalytic optimizing process, Technical Report TR91-016, Politecnico di Milano.

[10] M. Dorigo, G. Di Caro (1999), The Ant Colony Optimization Meta-Heuristic. In D. Corne, M. Dorigo, F. Glover (eds) New Ideas in Optimization, McGraw-Hill, London, UK, pp 11-32

[11] M. Dorigo, G. Di Caro, L.M. Gambardella (1999) Ant Algorithms for Discrete Optimization. Artificial Life, 5(2):137-172 25. M. Dorigo, L.M. Gambardella (1997) Ant colony system: a cooperative learning approach to the traveling salesman problem, IEEE Transaction on Evolutionary Computation 1:53-66

[12] Karaboga, D, (2005). An idea based on honey bee swarm for numerical optimization. Technical Report TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.

[13] Dervis Karaboga · Bahriye Basturk , A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, J Glob Optim (2007) 39:459–471

[14] Rajesh A. Thakker, M. Shojaei Baghini, and M. B. Patil Automatic Design of Low-Power Low-Voltage Analog Circuits Using Particle Swarm Optimization with Re-Initialization ,Department of Electrical Engineering, Indian Institute of Technology Bombay, 400076, India, Journal of Low Power Electronics Vol. 5, 1–12, 2009