An Improved Model for Ant based Clustering

Saroj Bala
AKG Engineering College,
Ghaziabad, U.P., India.

S. I. Ahson
Shobhit University,
Meerut, U.P., India.

R. P. Agarwal
Shobhit University,
Meerut, U.P., India.

ABSTRACT
Grouping different objects possessing inherent similarities in clusters has been addressed as the clustering problem among researchers. The development of new metaheuristics has given another direction to data clustering research. Swarm intelligence technique using ant colony optimization provides clustering solutions based on brood sorting. After basic ant model of clustering, number of improvements has been proposed. But the ant clustering still suffers with low convergence. This paper presents a novel model of intelligent movement of ants including the negative pheromone and direction selection. Negative pheromone plays a role of barrier in the direction of empty area and direction selection avoids the calculations not contributing to the clustering process. Simulations have shown good results.

Keywords
clustering, ant colony optimization, pheromone

1. INTRODUCTION
Nature inspired metaheuristics have become very popular among researchers. Ant Colony optimization (ACO) metaheuristic is inspired by collective behavior of ants that are doing complex tasks without any leader. In the early 1990s, ant colony optimization was introduced by M. Dorigo and colleagues for the solution of hard combinatorial optimization problems [2-4]. ACO is inspired by the foraging behavior of real ants. When searching for food, ants initially explore the area surrounding their nest in a random manner. If 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, a chemical pheromone trail is deposited on the ground by the ant. The quantity of pheromone deposited, which may depend on the quantity and quality of the food, will guide other ants to the food source. Stigmergy between the ants via pheromone trails enables them to find shortest paths between their nest and food sources. This characteristic of real ant colonies is exploited in artificial ant colonies in order to solve optimization problems.
Brood sorting in ant colonies has formed basis for ant based clustering algorithms. This paper explores some popular ant based algorithms for data clustering in section 2 and 3. An improved model for clustering is proposed in section 4. Section 5 covers experimental observations and results. Section 6 discusses conclusion.

2. BASIC ANTS
The basic ant clustering algorithm was proposed by Deneubourg [1]. In this model, the ants would walk randomly on the workspace, picking or dropping one data element from it. The ants possessed only local perceptual capabilities. They could sense the surrounding objects were similar or not to the object, they were carrying. Based on this information, they would perform the pick or drop action. The probability of picking and dropping an object depends on the objects lying in immediate environment. The probability of picking an object will be greater if less objects of same sort are there in the environment, as given by the following function:

$$P_{pick} = (\frac{K^+}{K^+ + f})^2$$

where f is an estimation of the fraction of nearby points occupied by objects of the same type, and K^+ is a constant. The probability decreases as f tends to l. The probability of dropping an object will be greater if more objects of the same sort are there in the immediate environment, as given by the following function:

$$P_{drop} = (\frac{f}{K^- + f})^2$$

where f is as same and K^- is a constant. The probability thus increases as f tends to 1.

3. RELATED WORK
The basic model was improved [5] by giving the ants the capacity to sense the complexity of their neighborhood. A complexity of ants surrounding area was determined by the presence or absence of objects, so that a place completely empty or completely full would have the lowest complexity, and a checkered pattern would have the highest. The tendency of the ants to take an action depends on the level of this complexity. The ants would not try to pick or drop anything in areas with low complexity. These complexity-seeking ants were thus able to avoid actions that did not contribute to the clustering process, performing their task more efficiently. A 9-cell neighborhood is considered at ant’s current position, which will have 12 internal faces. The complexity will be determined by the number of faces which separate the cells containing different type of objects and containing or not containing an object. Each ant calculates the local complexity. The ants can take a deterministic or a probabilistic approach. These ants spend less time in random movement in the area of low complexity and more time in careful processing at borders.
LF model is an extension of the basic ant model to cluster complex datasets into clusters [10]. The objects are able to differ among them along a continuous similarity measure. Each ant-like agent can only sense the similarity of the objects in their immediate region. The probability of picking or dropping an object was then a function of this measure of similarity. The ants move randomly on 2-d grid. For an unladen ant the probability of picking an object increases with low density and decreases with similarity of the object with the objects in a small surrounding area. The probability of picking an object i is defined as:

$$P_{pick}(i) = (\frac{k_p}{k_p + f(i)})^2$$

Where f(i) is an estimation of local density of elements and its similarity to I, defined as:

$$f(i) = \begin{cases} \frac{1}{d^2} \sum_j (1 - \frac{d(i,j)}{\alpha}, if (f > 0) \\ 0, otherwise \end{cases}$$

Where the probability of dropping an object is given by:

$$p_{drop}(i) = \begin{cases} 2f(i), f(i) < k_d \\ 1, otherwise \end{cases}$$

Where k_p and k_d are constants.

LF clustering model is applied to incremental web usage mining using active ants to take decision of movement and speed selection [7]. An improved formula is proposed for calculating the distance with connectivity, named as distance with connection [8]. The ant colony clustering algorithm based on the new distance calculation formula can discover clusters of arbitrary shapes. It is based on reachability paths between two objects. The similarity between x_i and x_j depends on the reachability paths between x_i and x_j that in turn depends on the connectivity. More reachability paths between x_i and x_j mean higher connectivity. Higher connectivity means less distance. Hence, distance with connection is inversely proportional to the connections.

A hybrid clustering algorithm is proposed based on density and ant colony algorithm [9]. It determines the initial cluster centers according to cluster objects distribution density method, and then uses the randomness of ant colony algorithm to find that arbitrary shape of clusters. Each object is assigned a density according to Gaussian function. A data set with n objects will be clustered into k classes; initial clusters will be those with highest density.

Jiang redefined the behavior of ant and colony similarity [6]. Initially only data is projected, the artificial ants are not scattered in the workspace. All the sample objects are randomly divided into several groups with the same number, and then are activated one by one to be artificial ants. The activated groups become a colony. The ants do not search for the sample objects scattered in the two-dimensional space as the artificial ants are the sample objects themselves. It can reduce the number of iterations and improve the efficiency of the algorithm. As the object samples around the activated artificial ant are also regarded as the ants, ant colony similarity is defined as:

$$fami(A_i) = \begin{cases} \sum_{O_j \in Neigh_i} (1 - \frac{diff(A_i, A_j)}{\alpha}, fami(A_i) > 0 \\ 0, otherwise \end{cases}$$

Where $Neigh_i$ represents the neighborhood of the artificial ant i. $Diff(A_i, A_j)$ represents the dissimilarity between vector A_i and vector A_j, which is measured by the distance between the two vectors. The distance is measured by the linear combination of Euclidean distance and cosine distance. The improved ant dissimilarity formula is defined by

$$diff(A_i, A_j) = k_1.d_{euclidean}(A_i, A_j) + k_2.d_{\cos}(A_i, A_j)$$

where k_1 and k_2 are distance coefficient constants, which represent respectively the weights of Euclidean distance and cosine distance on the measuring of the dissimilarity of the vectors, $k_1 \geq 0$, $k_2 \geq 0$, $k_1 + k_2 = 1$.

4. PROPOSED METHOD

Most of the ant based clustering methods are suffering from the low convergence. One reason behind is the random movement of the ant agents on the workspace (Here 2-D grid). If random movement makes the ant move in empty cell that may lead to explore the empty area that does not contribute to the clustering process. The proposed method is based on the intelligent movement of the ants that helps in picking and dropping the objects by the concept of direction selection and that avoids the exploration of empty area by the concept of negative pheromone.

4.1 Direction Selection

To drop an object the ant will move in the direction of that loaded object if present in surroundings. To pick an object the ant will move in the direction having minimum frequency in the surroundings. The surroundings are considered as the immediate 8 neighbor cells. A threshold is set, which is minimum number of surrounding objects, to take intelligent decision so that the initial iterations could run mainly according to ant memory. The process is summarized as shown:

Algorithm:

Step 1: Consider ant's current position and count elements in neighbor 8 cells.

Step 2: Do the following three steps if the number of elements in the surrounding cells is greater than the threshold otherwise ant moves randomly.

Step 3: Determine the frequency of different types of elements.

Step 4: An unloaded ant will move in the direction of objects with minimum frequency.

Step 5: A loaded ant will move in the direction of the object that matches with the load.

4.2 Negative Pheromone

It has been observed in existing ant clustering methods that the ants move in the empty area doing all the calculations but not resulting in clustering and sometimes making ants to recluster the objects. The proposed method handles it with help of a negative pheromone that acts as a barrier to stop ant movement in the direction of empty area. A negative pheromone is assigned to the empty cornered cells. The ants will not move in the direction of negative pheromone that will avoid the calculations in empty areas. But spread of negative pheromone in the major area of grid may disturb the process of clustering, so, negative pheromone is limited to grid corners and $1/4^{th}$ of the grid locations. Four exploring ants move from the four corners of the grid and lay negative pheromone on the empty corners of the grid. The ant movement is affected by this negative pheromone as shown:

Algorithm:

Step 1: Ant chooses a direction and calculates new location.

Step 2: If pheromone at new location is negative then the actual movement does not take place and a random direction is chosen.

Step 3: This process is repeated until an allowed location of movement is met.

5. EXPERIMENTAL RESULTS

The experiments are done with basic ant model with direction selection and negative pheromone. Ants have been simulated on a 2-D grid in MATLAB using GUI environment. Here 25 ants have been simulated with memory of size 10. The dropping and picking probabilities are taken same as basic ants. Parameters K^+ and K^- are taken as 0.1. Threshold for intelligent decision is taken as 2 and 4 to study the results. All the experiments are done on 30×30 grid. Negative pheromone is allowed up to $1/4^{th}$ of the grid. The pheromone is shown with help of green stars. Results are shown in Figure 1. Two different types of objects are taken, 100 objects of '○' type and 100 objects of '□ 'type. Figure 1(a) shows random distribution of objects and 1(b-f) shows the results of successive iterations.

6. CONCLUSION

Existing ant based clustering algorithms have been observed to show random movement of the ants that may cause the useless walk of ants in empty cells of the workspace. The proposed method is based on the intelligent movement of ants to avoid unnecessary computations. The movement has been made intelligent by putting negative pheromone on empty corners of the grid that avoids calculations in empty area and deciding the right direction of movement in workspace separately for loaded and unloaded ants. Experiments have shown improved results with lesser number of iterations and comparatively less calculations during one iteration.

7. REFERENCES

[1] Deneubourg, J. L., Goss, S., Franks, N., Franks, A., Detrain C. and Chretien, L. 1991. The dynamics of collective sorting: Robot-like ants and ant-like robots. Simulation of Adaptative Behavior: From Animals to Animats, pp. 356-363.

[2] Dorigo, M. and Socha, K. 2006. An Introduction to Ant Colony Optimization. Technical Report No. TR/IRIDIA/2006-010. Universite Libre de Bruxelles, Belgium.

[3] Dorigo, M., Maniezzo, V. and Colorni, A. 1996. Ant System: Optimization by a colony of cooperating agents. IEEE Trans. Syst., Man, and Cybern. Part B, 26(1), 29.

[4] Dorigo, M. and Gambardella. L.M. 1997. Ant Colony System: a cooperative learning approach to the traveling salesman problem. IEEE Trans. Evol. Comput., 1(1), 53.

[5] Gutowitz, H. 1993. Complexity-seeking ants. Proceedings of the Third European Conference on Artificial Life.

[6] Jiang, H., Yu Q. and Yu G. 2010. An Improved Ant Colony Clustering Algorithm. 3rd International Conference on Biomedical Engineering and Informatics published in IEEE 978-1-4244-6498-2/10. pp 2368-2372.

[7] Jie S., Ying L. and Zhimin, C. 2006. Incremental Web Usage Mining Based on Active Ant Colony Clustering. Wuhan Univ. Journal of Natural Sciences. pp. 1081-1085, Vol 11, No. 5.

[8] Li, S., Huang, W. and Tan, Y. 2010. An Improved Ant-Colony Clustering Algorithm Based on the Innovational Distance Calculation Formula. Third International Conference on Knowledge Discovery and Data Mining published in IEEE 978-0-7695-3923-2/10. pp. 342-346.

[9] Li, L., Wu W. and Rong Q. 2010. Research on Hybrid Clustering Based on Density and Ant Colony Algorithm. Second International Workshop on Education Technology and Computer Science published in IEEE 978-0-7695-3987-4/10, pp. 222-225.

[10] Lumer, E.D. and Faieta, B. 1994. Diversity and adaptation in populations of clustering ants. Proceedings of the Third International Conference on the Simulation of Adaptative Behavior: From Animals to Animats 3, pp. 449-508. MIT Press.

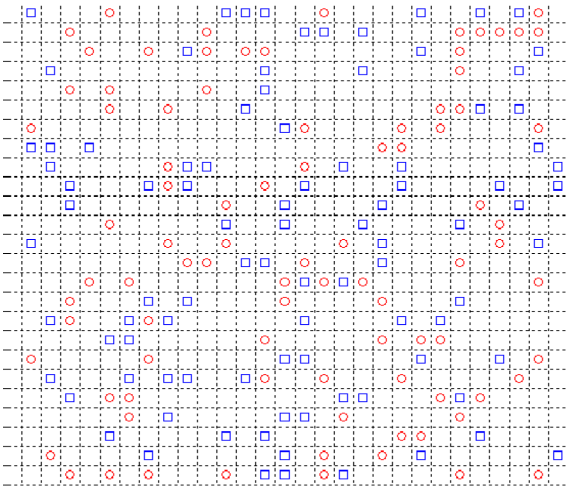

Figure 1(a) random distribution of objects on the grid

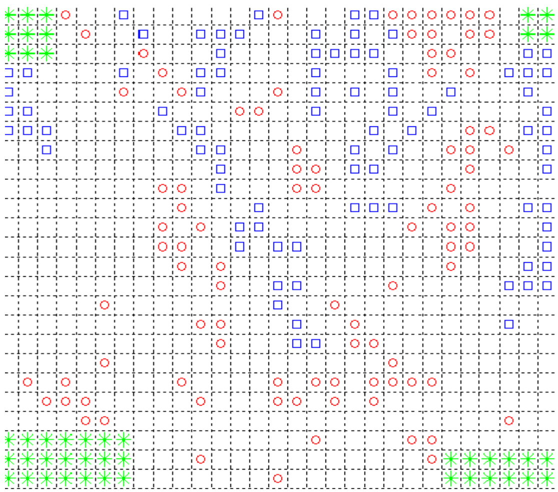

Figure 1(b) shows situation after 1000 iterations.

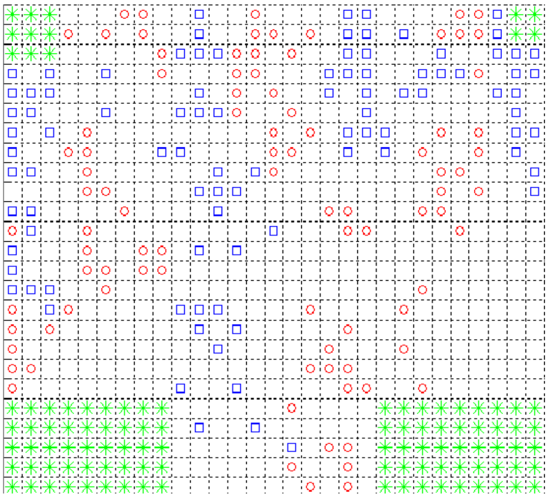

Figure 1(c) shows 2000 iterations

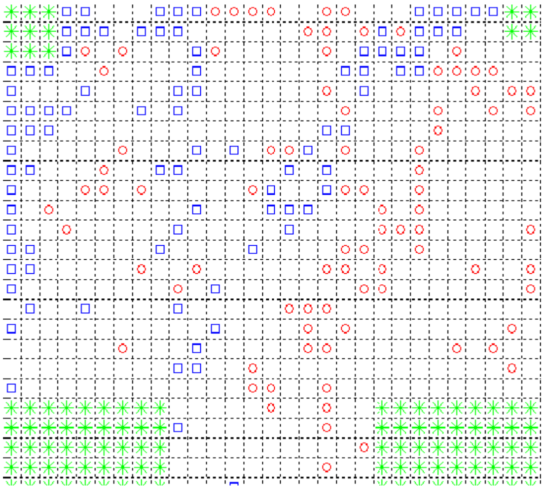

Figure 1(d) shows 3000 iterations

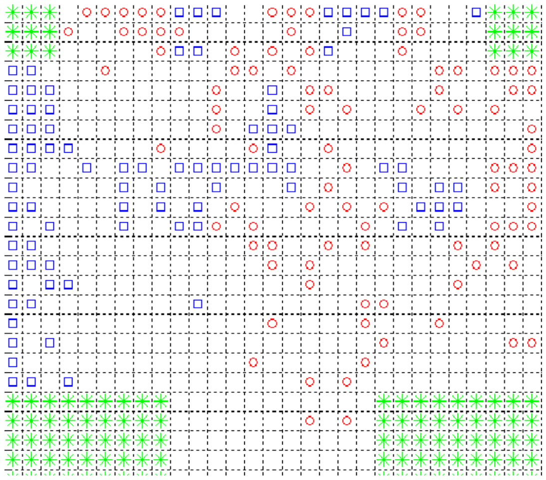

Figure 1(e) shows 4000 iterations

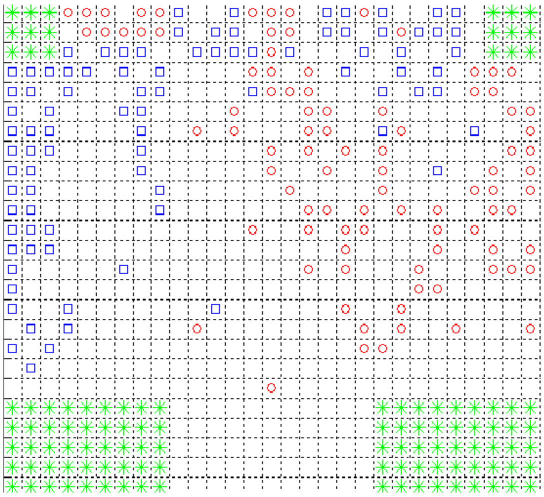

Figure 1(f) shows 5000 iterations

Figure 1 shows the simulated results